MORRIS, NICHOLS, ARSHT & TUNNELL LLP

1201 NORTH MARKET STREET
P.O. BOX 1347
WILMINGTON, DELAWARE 19899-1347

302 658 9200
302 658 3989 FAX

JACK B. BLUMENFELD
302.351.9291
jblumenfeld@mnat.com

Original Filing Date:
July 24, 2008

The Honorable Gregory M. Sleet
United States District Court
844 North King Street
Wilmington, DE 19801

BY ELECTRONIC FILING

REDACTED VERSION
Redacted Filing Date:
July 31, 2008

Re: Cisco v. Telcordia, No. 07-113-GMS

Dear Chief Judge Sleet:

In its opening summary judgment letter, Telcordia presents a number of arguments that it contends pose "fundamental, legal infringement issues" and supposedly establish that this case is "ripe for summary judgment." Telcordia's request ignores not only the evidence that was uncovered in discovery (and will be further explored when expert discovery commences), but the very foundations of summary judgment law. In so doing, Telcordia asks this Court to draw factual inference upon factual inference against Cisco and in favor of itself. Telcordia attempts to veil these inferences in caselaw citations and attorney argument to try to avoid the very real factual disputes that preclude summary judgment. Telcordia's request to move for summary judgment should be denied.

Telcordia Can Be Liable For Direct Infringement Of The Asserted Method Claims

Telcordia begins its letter with the contention that it cannot be liable for direct infringement because it "does not 'make' or 'sell'" the associated hardware that is necessary for its software to execute" and thus that "Telcordia's acts of 'making' and 'selling' its software cannot and do not amount to a direct infringement."

Notably absent from Telcordia's letter is any mention of Telcordia's '***use***' of the patented method, which constitutes direct infringement under Section 271(a). Even assuming for the sake of argument that Telcordia is correct that it never makes or sells the hardware on which its software runs and thus cannot infringe through those acts, Telcordia's own admissions confirm that Telcordia in fact *uses* the patented methods and thus very well can be liable for direct infringement. Specifically,

REDACTED

This testing is a use of the accused methods in the United States and itself constitutes a

direct infringement under 35 U.S.C. § 271(a). As a result, Telcordia's request for leave to seek summary judgment of no direct infringement should be rejected on this basis alone.

Material Factual Disputes Preclude Summary Judgment On Indirect Infringement

Telcordia's second basis for requesting summary judgment is as divorced from the factual record and principles of summary judgment as its first. This time, Telcordia's summary judgment theory is that it cannot be liable for indirect infringement because Cisco has not subpoenaed any Telcordia customers to identify a specific instance of direct infringement necessary for an indirect infringement finding. Telcordia argues that even if the accused methods fall within the scope of the claim, Cisco has not shown that anyone has practiced the accused methods. Once again, the law and facts omitted from Telcordia's letter are telling.

At the outset, Telcordia fails to acknowledge the legal principle that Telcordia itself trumpeted when it was the plaintiff in its case against Cisco, where the shoe was on the other foot. Just last year, in the *Telcordia v. Cisco* case, it was Telcordia that had opted not to subpoena *Cisco* customers for evidence of underlying instances of direct infringement and instead to proceed to trial on the basis of circumstantial evidence. When Cisco pointed this out, Telcordia was quick to note that "the law provides that the elements of a claim for indirect infringement may be shown through circumstantial evidence." Ex. C [Ex. I3.B to Final Pretrial Order in Case No. 04-876-GMS] at 25; *see also, e.g., Moleculon Research Corp. v. CBS, Inc.*, 793 F.2d 1261, 1272 (Fed. Cir. 1986) (holding that circumstantial evidence is adequate proof of direct infringement); *Lucent Techs., Inc. v. Gateway, Inc.*, 2008 WL 2491955 at *10-11 (S.D. Cal. June 19, 2008) (cataloging cases). Having relied on this law in its case against Cisco, Telcordia's failure to even acknowledge it here is particularly surprising.

As for the facts, Telcordia again ignores those that are most harmful to its request for summary judgment. Cisco will show both through factual evidence and through the parties' experts (expert discovery is upcoming) that

REDACTED

[1]

As one would expect based on this, Telcordia's actions resulted in precisely the instances of direct infringement it failed to acknowledge in its letter. As one example, Telcordia's Rule

---

1 Similarly, for the '622 Patent,

REDACTED

30(b)(6) witness on *inter alia* the "structure, characteristics and operation (actual or intended)" and "the installation, setup, maintenance and customer support" of the functionality accused of infringing the '988 Patent testified that

REDACTED

As Telcordia itself acknowledged in the prior case, circumstantial evidence of this nature can support an ultimate finding of direct infringement and, at a minimum, is more than sufficient to create a material factual dispute precluding summary judgment. *See, e.g., Arthrocare Corp. v. Smith & Nephew, Inc.*, 406 F.3d 1365, 1375-77 (Fed. Cir. 2005) (affirming denial of JMOL of no indirect infringement based in part on strong circumstantial evidence of direct infringement including product literature accompanying the product at issue); *Golden Blount, Inc. v. Robert H. Peterson Co.*, 438 F.3d 1354, 1363 (Fed. Cir. 2006).[2] In short, the threshold of circumstantial evidence of underlying direct infringement necessary to preclude summary judgment of no indirect infringement is far exceeded here. Whatever disputes remain as to direct infringement by Telcordia's customers, they remain for the jury to decide.

Material Factual Disputes As To Divided Infringement Preclude Summary Judgment

In its letter, Telcordia suggests that summary judgment of non-infringement is appropriate because infringement by the accused Telcordia software products is divided among multiple parties. As with its other arguments, Telcordia's argument misses the mark.

At the outset, the evidence discussed above supports a finding that Telcordia itself directly infringes the asserted method claims through, at least, the testing it runs in its own testing labs in New Jersey. In those instances, each step of the method claim would be performed on Telcordia's own network by Telcordia's own people.

Moreover, even in the cases where the direct infringement is by Telcordia's customers and not by Telcordia itself, Telcordia has not come close to showing (nor could it) that multiple parties are required for the underlying direct infringements. To the contrary, the asserted claims of the patents-in-suit both require that the method be performed by only a single Telcordia

---

[2] As the Court is all too familiar from the prior case, where the rubber really meets the road on the issue of underlying direct infringement is not on the threshold question of whether there is circumstantial or direct evidence of a single specific instance of direct infringement sufficient to preclude summary judgment of no indirect infringement (which there was in the prior case and is here), but whether there is evidence of the *quantum* of that underlying infringement as required for the damages analysis. The latter issue was at the heart of Cisco's motion *in limine* on damages in the prior case (to which *Golden Blount* was central), *see* D.I. 343 in Case No. 04-876-GMS, and will be an issue addressed by the parties' damages experts in this case. It is not an issue relevant to Telcordia's summary judgment letter.

customer (i.e., a single direct infringer). Even a cursory review of claim 1 of the '988 Patent reveals that a single Telcordia customer would perform each step of the asserted method claim. As an example, several of Telcordia's customers are telephone companies providing local telephone services. These customers would perform the accused method whenever a network manager generates, using the NMA system, a generic instruction that she wants to communicate to a group of network elements (e.g., telephone switches) within that telephone company's network and that instruction is then converted by the NMA system to specific instructions sent to those switches. Nowhere in that process is anyone other than a single Telcordia customer involved in the direct infringement underlying Telcordia's indirect infringement.

Similarly, in the case of the '622 Patent, it is Telcordia's customers who each, on their own, directly infringe the asserted method claim. Telcordia's suggestion to the contrary is based on a misapprehension of not only the claim language but of Cisco's infringement contentions. To be sure, asserted claim 7 requires that an application program is executed on a first and second processing system. But these systems may well be within the same customer's network. Indeed, Telcordia appears to fundamentally misunderstand Cisco's infringement contentions when it argues that steps of the asserted method claim are met by Cisco's own network devices operating Cisco's own software. In fact, it is Telcordia's customers who deploy Cisco's own network devices (as an example) within their networks, and by executing applications on Cisco's or other equipment manufacturers' devices operating within their own networks, meet the claim requirement.

It is perhaps understandable that Telcordia's motion would be based on these misunderstandings of the asserted claims and Cisco's infringement contentions because expert discovery has yet to even commence. Once it is underway, expert discovery will only serve to further confirm that Cisco's infringement allegations go to direct infringement of the asserted method claims by Telcordia's customers, *one at a time.* There is at least a material dispute of fact that precludes summary judgment on the question of divided infringement.

A Material Factual Dispute Precludes Summary Judgment On Claim 7 Of The '622 Patent

Telcordia's fourth proposed ground for summary judgment is based on this Court's construction of "socket" to mean "an application program interface (API) that was *developed for the Berkeley version of AT&T's UNIX operating system ....*" As Cisco and surely the Court anticipated, Telcordia has latched on to this claim construction to support a non-infringement argument that the accused products do not operate on a system developed for the Berkeley version of AT&T's UNIX operating system. At first blush, Telcordia's argument seems to warrant further consideration. The problem with this is, of course, that summary judgment of non-infringement cannot be based on a superficial comparison of the claims to the accused products devoid of an evidentiary analysis or the testimony of experts on how the claim limitations may or may not be met by the accused products.

Here, that problem is particularly acute because the record shows that Telcordia's non-infringement theory is far from clean-cut. To the contrary, the evidence shows that this claim requirement is in fact met by the accused products, or at a minimum, would be met under the

doctrine of equivalents. As but one example, a *BSD Sockets Interface Programmer's Guide* on HP's UX (UNIX-based) operating system explains that the sockets used by that system are in fact based on the sockets developed at Berkeley and function in much the same way. *See, e.g.,* Ex. E [BSD Sockets Interface Programmer's Guide] at CISCO-0000122 ("HP's implementation of BSD Sockets is a full set of sockets from the networking services originally developed by the University of California at Berkeley (UCB)."). Expert testimony as to any insubstantial differences between the accused products and an API developed for the Berkeley version of AT&T's UNIX operating system will serve to further illuminate the infringement debate. As such, Telcordia's motion for summary judgment as to infringement under these circumstances would flip summary judgment on its head and require the Court to draw every inference in Telcordia's favor. Telcordia's request that the Court do so should be rejected.

A Material Factual Dispute Precludes Summary Judgment On Claim 1 Of The '988 Patent

As with the '622 Patent, the record supports inferences that are quite the opposite from those Telcordia relies upon for summary judgment on the '988 Patent.

For the '988 Patent, Telcordia contends that the claim requirement "generic instruction," which the parties agreed means "an instruction applicable to the groups of elements," is not met because the accused system sends instructions to network elements one at a time. Telcordia's non-infringement ignores two fundamental facts confirming that the accused systems do in fact send "generic instructions" as required by the claims. First, Telcordia ignores the fact that its own Rule 30(b)(6) witnesses on the accused products admitted numerous times in deposition that

REDACTED

Second, Telcordia ignores the fact that the invention of the '988 Patent, as it was conceived, reduced to practice, described and claimed, requires that a generic command be translated into device-specific commands that are sent one at a time. According to an inventor of the '988 patent, this is precisely what he and his co-inventor envisioned and implemented and, as a technical matter, is the only way a system like this could work. *See, e.g.,* Ex. F [6/27/08 Spector Dep.] at 109:17-112:25; 213:1-18. Both of these facts are supported in the factual record that Telcordia ignores and both will surely be the subject of extensive expert proof.

Once again, the only way Telcordia's request for summary judgment can be granted is if all factual inferences are drawn in Telcordia's favor and expert discovery on fundamental infringement issues is completely forgone. Neither the law nor the facts would support such a request.

* * *

For the reasons set forth above, Telcordia's request to file a motion for summary judgment of non-infringement should be denied.

Respectfully submitted,

/s/ Jack B. Blumenfeld

Jack B. Blumenfeld (#1014)

cc: Peter T. Dalleo, Clerk (By Hand)
Edward R. Reines, Esquire (By Email)
Steven J. Balick, Esquire (By Hand and Email)
Vincent P. Kovalick, Esquire (By Email)

EXHIBIT A

FULLY REDACTED

EXHIBIT B

FULLY REDACTED

EXHIBIT C

FULLY REDACTED

EXHIBIT D

FULLY REDACTED

EXHIBIT E

BSD Sockets Interface Programmer's Guide

Edition 6

B2355-90136
HP 9000 Networking
E0497

Printed in: United States

Legal Notices

The information in this document is subject to change without notice.

Hewlett-Packard makes no warranty of any kind with regard to this manual, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. Hewlett-Packard shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

Warranty. A copy of the specific warranty terms applicable to your Hewlett- Packard product and replacement parts can be obtained from your local Sales and Service Office.

Restricted Rights Legend. Use, duplication or disclosure by the U.S. Government is subject to restrictions as set forth in subparagraph (c) (1) (ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.227-7013 for DOD agencies, and subparagraphs (c) (1) and (c) (2) of the Commercial Computer Software Restricted Rights clause at FAR 52.227-19 for other agencies.

HEWLETT-PACKARD COMPANY 3000 Hanover Street Palo Alto, California 94304 U.S.A.

Use of this manual and flexible disk(s) or tape cartridge(s) supplied for this pack is restricted to this product only. Additional copies of the programs may be made for security and back-up purposes only. Resale of the programs in their present form or with alterations, is expressly prohibited.

This software is based in part on the Fourth Berkeley Software Distribution under license from the Regents of the University of California.

Contents

Contents

Contents

Contents

Contents

Contents

Printing History

The manual printing date and part number indicate its current edition. The printing date will change when a new edition is printed. Minor changes may be made at reprint without changing the printing date. the manual part number will change when extensive changes are made.

Manual updates may be issued between editions to correct errors or document product changes. To ensure that you receive the updated or new editions, you should subscribe to the appropriate product support service. See your HP sales representative for details.

First Edition: February 1991

Second Edition: July 1992 (HP-UX Release 9.0)

Third Edition: January 1995 (HP-UX Release 10.0)

Fourth Edition: December 1995 (HP-UX Release 10.10)

Fifth Edition: June 1996

Sixth Edition: April 1997 (HP-UX Release 10.30)

Preface

This guide describes HP's BSD Sockets. BSD Sockets is a set of programming development tools for interprocess communication. HP's implementation of BSD Sockets is a full set of sockets from the networking services originally developed by the University of California at Berkeley (UCB).

The *BSD Sockets Interface Programmer's Guide* is the primary reference for programmers who write or maintain BSD Sockets applications on HP 9000 computers. This guide also provides information about BSD Sockets features for node managers who design HP 9000 networks.

This guide is written for intermediate to advanced programmers and assumes that:

- Your HP 9000 networking is installed and running on your local host.
- You have all the login names you may be associated with (these may be obtained from your node manager).
- You have a list of other hosts or nodes your HP 9000 product can communicate with (this list may also be obtained from your node manager).

This guide is organized as follows:

Chapter 1	**BSD Sockets Concepts** provides an overview of the Client-Server Model.
Chapter 2	**Using Internet Stream Sockets** describes the steps involved in creating an internet stream socket using the AF_INET address family.
Chapter 3	**Advanced Topics for Stream Sockets** explains socket options, synchronous I/O multiplexing with `select`, sending and receiving data asynchronously, nonblocking I/O, using `shutdown`, using `read` and `write` to make stream sockets transparent, and sending and receiving out-of- band data.
Chapter 4	**Using Internet Datagram Sockets** describes the steps involved in creating an internet datagram socket connection.

Chapter 5	**Advanced Topics for Internet Datagram Sockets** includes information about default socket address, synchronous I/O multiplexing with `select`, sending and receiving IP multicast datagrams, and broadcast addresses.
Chapter 6	**Using UNIX Domain Stream Sockets** describes the steps involved in creating a UNIX Domain stream socket connection between two processes executing on the same node.
Chapter 7	**Using UNIX Domain Datagram Sockets** describes the steps required to exchange data between AF_UNIX datagram sockets on the same node without establishing a connection.
Chapter 8	**Programming Hints** contains information about troubleshooting, port addresses, diagnostic utilities, internet daemon, `inetd`, portability issues, and system and library calls.
Appendix A	**BSD Sockets Quick Reference Table** summarizes the differences in calls for the two sets of sockets behavior: HP-UX BSD Sockets and X/Open Sockets.

1 BSD Sockets Concepts

The BSD Sockets facility allows you to create distributed applications that pass data between programs (on the same computer or on separate computers on the network) without requiring an understanding of the

many layers of networking protocols. This is accomplished by using a set of system calls. These system calls allow you to create communication endpoints called **sockets** and transfer data between them.

NOTE BSD Sockets is a program development tool. Before you attempt to use BSD Sockets, you may need to familiarize yourself with the C programming language and the HP-UX operating system. You could implement a BSD Sockets application using FORTRAN or Pascal, but all library calls and include files are implemented in C.

Introduction

This guide describes the steps involved in establishing and using BSD Sockets connections. It also describes the protocols you must use and how the BSD Sockets system calls interact. The details of each system call are described in the corresponding man pages.

Key Terms and Concepts

For a basic understanding of BSD Sockets and its general model, you should review the following terms and definitions.

address family The address format used to interpret addresses specified in socket operations. The internet address family (AF_INET) and the Berkeley UNIX Domain address family (AF_UNIX) are supported.

addressing A means of labeling a socket so that it is distinguishable from other sockets on a host.

association A BSD Sockets connection is defined by an association. An AF_INET association contains the (protocol, local address, local port, remote address, remote port)-tuple. An AF_UNIX association contains the (protocol, local address, peer address)-tuple. Associations must be *unique*; duplicate associations on the same host cannot exist. The tuple is created when the local and remote socket addresses are bound and connected. This means that the association is created in two steps, and there is a chance that two potential associations could be alike between steps. The host prevents duplicate associations by checking for uniqueness of the tuple at connection time, and reporting an error if the tuple is not unique.

binding	Before a socket can be accessed across the network, it must be bound to an address. Binding associates a socket address with a socket and makes the socket accessible to other sockets on the network. Once a socket address is bound, other sockets can connect to the socket and send data to or receive data from it.
channel	Communication path created by establishing a connection between sockets.
communication domain	A set of properties that describes the characteristics of processes communicating through sockets. The internet (AF_INET) address family domain is supported. The UNIX Domain (AF_UNIX) address family domain is also supported, for local communication only.
internet address	A four-byte address that identifies a node on the network.
message	The data sent in one UDP packet.
packet	A message or data unit that is transmitted between communicating processes.
peer	The remote process with which a process communicates.
port	An address within a host that is used to differentiate between multiple sockets with the same internet address. You can use port address values 1024 through 65535. (Port addresses 1 through 1023 are reserved for the super-user.)
protocols	Two internet transport layer protocols can be used with BSD Sockets. They are TCP, which implements stream sockets, and UDP, which implements datagram sockets.
socket	Sockets are communication endpoints. A pair of connected sockets provides an interface similar to that of HP-UX pipes. A socket is identified by a socket descriptor.

socket address	For the internet address family (AF_INET), the socket address consists of the internet address, port address and address family of a socket. The internet and port address combination allows the network to locate a socket. For UNIX Domain (AF_UNIX), the socket address is the directory pathname bound to the socket.
socket descriptor	A socket descriptor is an HP-UX file descriptor that references a socket instead of an ordinary file. Therefore, it can be used for reading, writing, or most standard file system calls after a BSD Sockets connection is established. System calls that use file descriptors (e.g. `read`, `write`, `select`) can be used with socket descriptors. All BSD Sockets functions use socket descriptors as arguments.
TCP	Provides the underlying communication support for stream sockets. The Transmission Control Protocol (TCP) is used to implement reliable, sequenced, flow-controlled two-way communication based on byte streams similar to pipes. Refer to the `tcp(7p)` man page for more information on TCP.
UDP	Provides the underlying communication support for datagram sockets. The User Datagram Protocol (UDP) is an unreliable protocol. A process receiving messages on a datagram socket could find messages are duplicated, out-of-sequence, or missing. Messages retain their record boundaries and are sent as individually addressed packets. There is no concept of a connection between the communicating sockets. Refer to the `udp(7p)` man page for more information on UDP.
UNIX Domain Protocol	In addition, the UNIX Domain protocol may be used with AF_UNIX sockets for interprocess communication on the same node. Refer to the `unix(7p)` man page for more information on the UNIX Domain protocol.

How You Can Use BSD Sockets

The best example of how BSD Sockets can be used is the Internet Services. These services use BSD Sockets to communicate between remote hosts. Using the BSD Sockets facility, you can write your own distributed application programs to do a variety of tasks.

For example, you can write distributed application programs to:

- Access a remote database.
- Access multiple computers at one time.
- Spread tasks across several hosts.

The Client-Server Model

Typical BSD Sockets applications consist of two separate application level processes; one process (the **client**) requests a connection and the other process (the **server**) accepts it.

The server process creates a socket, binds an address to it, and sets up a mechanism (called a listen queue) for receiving connection requests. The client process creates a socket and requests a connection to the server process. Once the server process accepts a client process's request and establishes a connection, full-duplex (two-way) communication can occur between the two sockets.

This set of conventions must be implemented by both processes. Depending upon the needs of your application, your implementation of the model can be symmetric or asymmetric. In a symmetrical application of the model, either process can be a server or a client. In an asymmetrical application of the model, there is a clearly defined server process and client process. An example of an asymmetrical application is the `ftp` service.

Creating a Connection: the Client-Server Model

The following figures illustrate conceptual views of the client-server model at three different stages of establishing a connection. The completed steps are included in each figure.

Figure 1-1 **Client-Server in a Pre-Connection State**

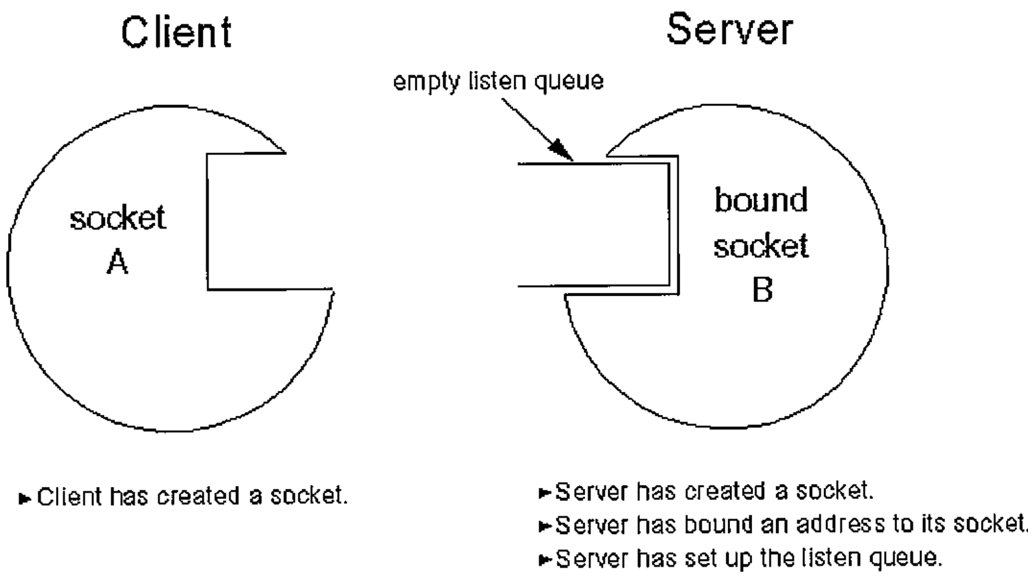
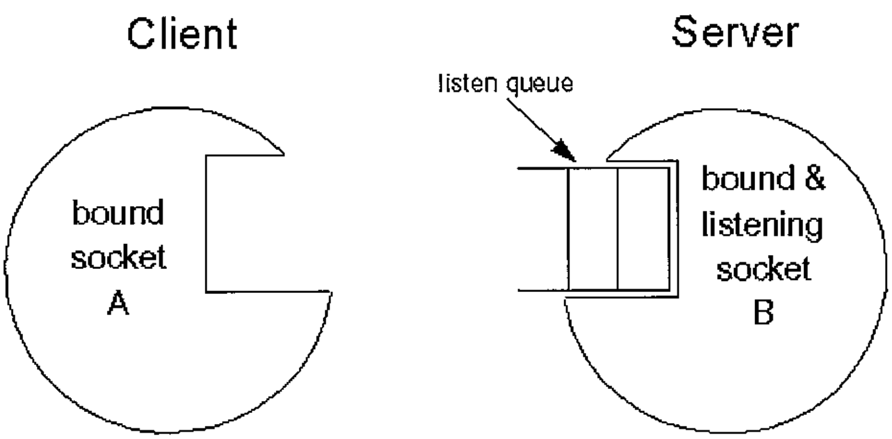


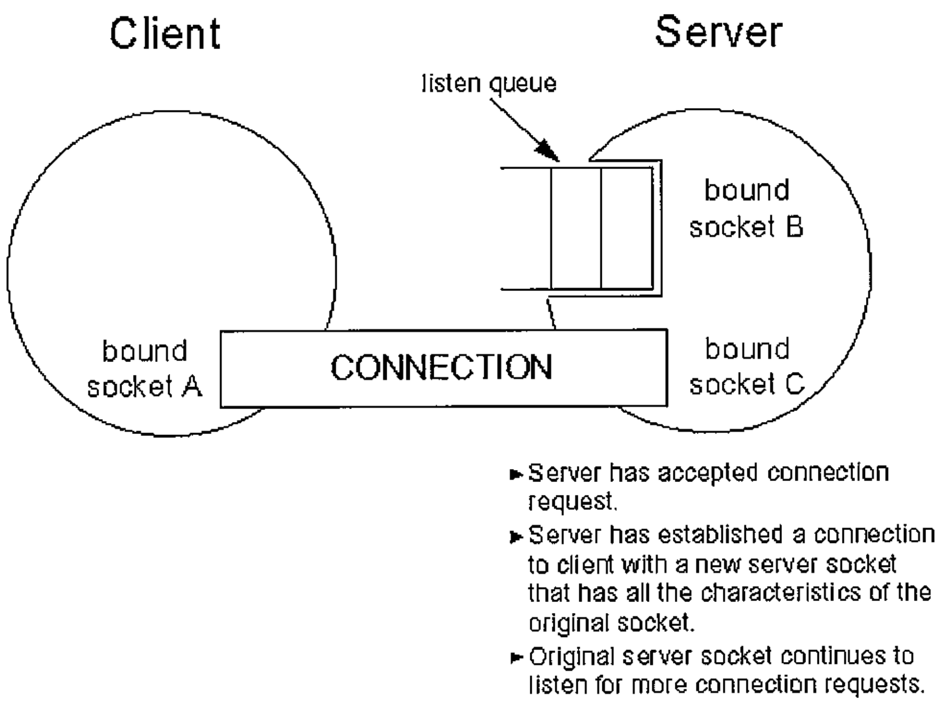
Figure 1-2 **Client-Server at Time of Connection Request**



▶ Client has made a connection request.

▶ Server has received the request in the listen queue.

Figure 1-3 **Client-Server When Connection is Established**



BSD Sockets Library Routines

The library routines and system calls that you need to implement a BSD Sockets application are described throughout the guide. In addition, a complete list of all these routines and system calls is provided in the "Summary Tables for Library and System Calls" section of chapter 8, "Programming Hints."

The library routines are in the common "c" library named `libc.sl`. Therefore, you do not need to specify any library name on the cc command line to use these library calls, `libc.sl` is used automatically.

2 Using Internet Stream Sockets

NOTE

Release 10.10 and later releases support two variations of sockets behavior: classic HP-UX sockets and X/Open sockets. By default, users receive classic HP-UX sockets. To use X/Open sockets, users must make an addition to their make files by including the "-l xnet" argument with the "c89" or "cc" utilities.

Users may also use the "lint" utility to check for possible violations to the X/Open sockets specification by including "-l xnet" to the argument list of the lint utility.

Users must also define the following for X/Open sockets behavior:

- _XOPEN_SOURCE;
- _XOPEN_SOURCE_EXTENDED = 1

See Appendix A for a table which summarizes the differences in calls for the two variations of sockets behavior. A quick reference card listing these call differences is also available. For detailed information about HP-UX sockets, see the remainder of this manual as well as the man pages. For detailed information about X/Open sockets, see "CAE Specification, Networking Services, Issue 4, X/Open" and the man pages.

This chapter describes creating an internet stream socket connection using the AF_INET address family.

Overview

Internet TCP stream sockets provide bidirectional, reliable, sequenced and unduplicated flow of data without record boundaries. The following table lists the steps involved in creating and terminating a BSD Sockets connection using stream sockets.

Table 2-1 Creating/Terminating BSD Sockets Connections Using Internet Stream Sockets

Client Process Activity	System Call Used	Server Process Activity	System Call Used
create a socket	`socket()`	create a socket	`socket()`
bind a socket address (optional)	`bind()`	bind a socket address	`bind()`
		listen for incoming connection requests	`listen()`
request a connection	`connect()`		
		accept connection	`accept()`
send data	`write()` or `send()`		
		receive data	`read()` or `recv()`
		send data	`write()` or `send()`
receive data	`read()` or `recv()`		
disconnect socket (optional)	`shutdown()` or `close()`	disconnect socket (optional)	`shutdown()` or `close()`

Each of these steps or activities is described in more detail in the following sections. The description of each activity specifies a system call and includes:

- What happens when the system call is used.
- When to make the call.
- What the parameters do.
- How the call interacts with other BSD Sockets system calls.
- Where to find details on the system call.

The stream socket program examples are at the end of these descriptive sections. You can refer to the example code as you work through the descriptions.

Preparing Address Variables

Before you create a connection, establish the correct variables and collect the information that you need to request a connection.

Your server process needs to:

- Declare socket address variables.
- Assign a wildcard address.
- Get the port address of the service that you want to provide.

Your client process needs to:

- Declare socket address variables.
- Get the remote host's internet address.
- Get the port address for the service that you want to use.

These activities are described next. Refer to the program example at the end of this chapter to see how these activities work together.

Declaring Socket Address Variables

You need to declare a variable of type struct `sockaddr_in` to use for socket addresses. For example, the following declarations are used in the example client program:

```
struct sockaddr_in myaddr; /* for local socket address  */
struct sockaddr_in peeraddr; /* for peer socket address  */
```

`sockaddr_in` is a special case of `sockaddr` and is used with the AF_INET addressing domain. Both types are shown in this chapter, but `sockaddr_in` makes it easier to manipulate the internet and port addresses. Some of the BSD Sockets system calls are declared using a pointer to `sockaddr`, but you can also use a pointer to `sockaddr_in`.

The `sockaddr_in` address structure consists of the following fields:

Field	Description
`short sin_family`	Specifies the address family and should always be set to AF_INET.
`u_short sin_port`	Specifies the port address. Assign this field when you bind the port address for the socket or when you get a port address for a specific service.
`struct inaddr sin_addr`	Specifies the internet address. Assign this field when you get the internet address for the remote host.

The server process only needs an address for its own socket. Your client process may not need an address for its local socket. Refer to the `inet(7f)` man page for more information on `sockaddr_in`.

Getting the Remote Host's Internet Address

`gethostbyname` obtains the internet address of the host and the length of that address (as the size of `struct in_addr`) from `/etc/hosts` or from NIS name server. `gethostbyname` and its parameters are described in the following table.

Include files: `#include <netdb.h>`

System call:

```
struct hostent *gethostbyname(name)
char *name;
```

Parameter	Description of Contents	INPUT Value
name	pointer to a valid host name (null-terminated string)	host name string

Function result: pointer to struct `hostent` containing internet address
NULL pointer (0) if failure occurs.

Example:

```
#include <netdb.h>
struct hostent *hp; /* pointer to host info for remote host */
...
peeraddr.sin_family = AF_INET;
hp = gethostbyname (argv[1]);
peeraddr_in.sin_addr.s_addr = ((struct in_addr *)(hp->h_addr))->s_addr;
```

The `argv[1]` parameter is the host name specified in the client program command line. Refer to the `gethostent(3n)` man page for more information on `gethostbyname`.

Getting the Port Address for the Desired Service

When a server process is preparing to offer a service, it must get the port address for the service from `/etc/services` so it can bind that address to its "listen" socket. If the service is not already in `/etc/services`, you must add it.

When a client process needs to use a service that is offered by some server process, it must request a connection to that server process's "listening" socket. The client process must know the port address for that socket.

`getservbyname` obtains the port address of the specified service from `/etc/services`. `getservbyname` and its parameters are described in the following table.

Include files: `#include <netdb.h>`

System call:

```
struct servent *getservbyname(name, proto)
char *name, *proto;
```

Parameter	Description of Contents	INPUT Value
name	pointer to a valid service name	service name
proto	pointer to the protocol to be used	`tcp` or 0 if TCP is the only protocol for the service

Function result: pointer to struct `servent` containing port address NULL, pointer (0) if failure occurs.

Example:

```
#include <netdb.h>
struct servent *sp; /* pointer to service info */
...
sp = getservbyname ("example", "tcp");
peeraddr.sin_port = sp->s_port;
```

When to Get the Server's Socket Address

The server process needs to get the socket address before binding the listen socket. The client process needs to get the socket address before the client executes a connection request. Refer to the getservent(3n) man page for more information on getservbyname.

Using a Wildcard Local Address

Wildcard addressing simplifies local address binding. When an address is assigned the value of INADDR_ANY, the host interprets the address as any valid address. This is useful for your server process when you are setting up the listen socket. It means that the server process does not have to look up its own internet address. When INADDR_ANY is used as a host address, it also allows the server to listen to all network connections on the host. When a specific address is used in the bind, the server can only listen to that specific connection. Thus, INADDR_ANY is useful on a system in which multiple LAN cards are available, and messages for a given socket can come in on any of them.

For example, to bind a specific port address to a socket, but leave the local internet address unspecified, the following source code could be used:

```
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
...
struct sockaddr_in sin;
...
s = socket(AF_INET, SOCK_STREAM, 0);
sin.sin_family = AF_INET;
sin.sin_addr.s_addr = INADDR_ANY;
sin.sin_port = MYPORT;
bind (s, &sin, sizeof(sin));
```

Writing the Server Process

This section explains the calls your server process must make to connect with and serve a client process.

Creating a Socket

The server process must call `socket` to create a communication endpoint. `socket` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:
```
s = socket(af, type, protocol)
int af, type, protocol;
```

Parameter	Description of Contents	INPUT Value
af	address family	AF_INET
type	socket type	SOCK_STREAM
protocol	underlying protocol to be used	0 (default) or value returned by `getprotobyname`

Function result: socket number (HP-UX file descriptor), –1 if failure occurs.

Example:

```
s = socket (AF_INET, SOCK_STREAM, 0);
```

The socket number returned is the socket descriptor for the newly created socket. This number is an HP-UX file descriptor and can be used for reading, writing or any standard file system calls after a BSD Sockets connection is established. A socket descriptor is treated like a file descriptor for an open file.

When to Create Sockets

The server process should create a socket before any other BSD Sockets system calls. Refer to the socket(2) man page for more information on socket.

Binding a Socket Address to the Server Process's Socket

After your server process has created a socket, it must call bind to bind a socket address. Until an address is bound to the server socket, other processes have no way to reference it.

The server process must bind a specific port address to this socket, which is used for listening. Otherwise, a client process would not know what port to connect to for the desired service.

Set up the address structure with a local address before you make a bind call. Use a wildcard address so your server process does not have to look up its own internet address. bind and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <netinet/in.h>
#include <sys/socket.h>
```

System call:

```
bind (s, addr, addrlen)
int s;
struct sockaddr *addr;
int addrlen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket to be bound
addr	socket address	pointer to address to be bound to s
addrlen	length of socket address	size of struct sockaddr_in

Function result: 0 if bind is successful, –1 if failure occurs.

Example:

```
struct sockaddr_in myaddr;
...
bind (ls, &myaddr, sizeof(struct sockaddr_in));
```

When to Bind Socket Addresses

The server process should bind the socket address after the socket is created and before any other BSD Sockets system calls. Refer to the `bind(2)` man page for more information on `bind`.

Setting Up the Server to Wait for Connection Requests

Once your server process has an address bound to it, it must call `listen` to set up a queue that accepts incoming connection requests. The server process then monitors the queue for requests (using `select(2)` or `accept`). The server process cannot respond to a connection request until it has executed `listen`. `listen` and its parameters are described in the following table.

Include files: none

System call:

```
listen(s, backlog)
int s, backlog;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	server socket's descriptor
backlog	preferred maximum number of connection requests in the queue at any time	size of queue (between 0 and SOMAXCONN)

Function result: 0 if listen is successful, –1 if failure occurs.

Example:

```
listen (ls, 5);
```

`backlog` is the preferred number of unaccepted incoming connections allowed at a given time. The actual number may be greater than the specified backlog. When the request is full, further connection requests are rejected.

A backlog of 0 specifies only 1 pending connection can exist at any given time. SOMAXCONN is defined in `<sys/socket.h>`. The default setting is 20.

When to Set Up Server to Listen

The server process should be set up to listen after socket is created and bound and before the server can respond to connection requests. Refer to the `listen(2)` man page for more information on `listen`.

Accepting a Connection

The server process can accept any connection requests that enter its queue after it executes `listen`. `accept` creates a new socket for the connection and returns the socket descriptor for the new socket. The new socket:

- Is created with the same properties as the old socket.
- Has the same bound port address as the old socket.
- Is connected to the client process's socket.

`accept` blocks until there is a connection request from a client process in the queue, unless you are using nonblocking I/O. `accept` and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <netinet/in.h>
#include <sys/socket.h>
```

System call:

```
s = accept(ls,addr,addrlen)
int s;
int ls;
struct sockaddr *addr;
int *addrlen;
```

Parameter	Contents	INPUT Value	OUTPUT Value
ls	socket descriptor of local socket	socket descriptor of server socket	unchanged
addr	socket address	pointer to address structure where address will be put	pointer to socket address of client socket that server's new socket is connected to
addrlen	length of address	pointer to the size of struct `sockaddr_in`	pointer to the actual length of address returned in `addr`

Function result: socket descriptor of new socket if accept is successful, –1 if failure occurs.

Example:

```
struct sockaddr_in peeraddr;
...
addrlen = sizeof(sockaddr_in);
s = accept (ls, &peeraddr, &addrlen);
```

There is no way for the server process to indicate which requests it can accept. It must accept all requests or none. Your server process can keep track of which process a connection request is from by examining the address returned by `accept`. Once you have this address, you can use `gethostbyaddr` to get the hostname. You can close down the connection if you do not want the server process to communicate with that particular client host or port.

When to Accept a Connection

The server process should accept a connection after executing the `listen` call. Refer to the `accept(2)` man page for more information on `accept`.

Writing the Client Process

This section explains the calls your client process must make to connect with and be served by a server process.

Creating a Socket

The client process must call `socket` to create a communication endpoint. `socket` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:
```
s = socket(af, type, protocol)
int af, type, protocol;
```

Parameter	Description of Contents	INPUT Value
af	address family	AF_INET
type	socket type	SOCK_STREAM
protocol	underlying protocol to be used	0 (default) or value returned by `getprotobyname`

Function result: socket number (HP-UX file descriptor), –1 if failure occurs.

Example:

```
s = socket (AF_INET, SOCK_STREAM, 0);
```

The socket number returned is the socket descriptor for the newly created socket. This number is an HP-UX file descriptor and can be used for reading, writing or any standard file system calls after a BSD Sockets connection is established. A socket descriptor is treated like a file descriptor for an open file.

When to Create Sockets

The client process should create sockets before requesting a connection. Refer to the `socket(2)` man page for more information on `socket`.

Requesting a Connection

Once the server process is listening for connection requests, the client process can request a connection with the `connect` call. `connect` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <netinet/in.h>
#include <sys/socket.h>
```

System call:
```
connect(s, addr, addrlen)
int s;
struct sockaddr *addr;
int addrlen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket requesting connection
addr	pointer to the socket address	pointer to the socket address of the socket to which client wants to connect
addrlen	length of address	size of address structure pointed to by addr

Function result: 0 if connect is successful, –1 if failure occurs.

Example:

```
struct sockaddr_in peeraddr;
...
connect (s, &peeraddr, sizeof(struct sockaddr_in));
```

`connect` initiates a connection and blocks if the connection is not ready, unless you are using nonblocking I/O. When the connection is ready, the client process completes its `connect` call and the server process can complete its `accept` call.

NOTE

The client process does not get feedback that the server process has completed the `accept` call. As soon as the `connect` call returns, the client process can send data. Local internet and port addresses are bound when `connect` is executed if you have not already bound them yourself. These address values are chosen by the local host.

When to Request a Connection

The client process should request a connection after socket is created and after server socket has a listening socket. Refer to the `connect(2)` man page for more information on `connect`.

Sending and Receiving Data

After the connect and accept calls are successfully executed, the connection is established and data can be sent and received between the two socket endpoints. Because the stream socket descriptors correspond to HP-UX file descriptors, you can use the read and write calls (in addition to recv and send) to pass data through a socket-terminated channel.

If you are considering the use of the read and write system calls instead of the send and recv calls described below, you should consider the following:

- If you use read and write instead of send and recv, you can use a socket for stdin or stdout.
- If you use read and write instead of send and recv, you cannot use the options specified with the send or recv flags parameter.

See the table that lists other system calls in chapter 8, for more information on which of these system calls are best for your application.

Sending Data

send and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:

```
count = send(s,msg,len,flags)
int s;
char *msg;
int len, flags;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket sending data

Parameter	Description of Contents	INPUT Value
msg	pointer to data buffer	pointer to data to be sent
len	size of data buffer	size of msg
flags	settings for optional flags	0 or MSG_OOB

Function result: number of bytes actually sent, –1 if failure occurs.

Example:

```
count = send (s, buf, 10, 0);
```

`send` blocks until the specified number of bytes have been queued to be sent, unless you are using nonblocking I/O.

When to Send Data

The server or client process should send data after the connection is established. Refer to the `send(2)` man page for more information on `send`.

Receiving Data

`recv` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:
```
count = recv(s,buf,len,flags)
int s;
char *buf;
int len, flags;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket receiving data
buf	pointer to data buffer	pointer to buffer that is to receive data
len	maximum number of bytes that should be received	size of data buffer
flags	settings for optional flags	0, MSG_OOB or MSG_PEEK

Function result: number of bytes actually received, –1 if failure occurs.

Example:

```
count = recv(s, buf, 10, 0);
```

`recv` blocks until there is at least 1 byte of data to be received, unless you are using nonblocking I/O. The host does not wait for `len` bytes to be available; if less than `len` bytes are available, that number of bytes are received.

No more than `len` bytes of data are received. If there are more than `len` bytes of data on the socket, the remaining bytes are received on the next `recv`.

Flag Options

The flag options are:

- 0 for no options.
- MSG_OOB for out of band data.
- MSG_PEEK for a nondestructive read .

Use the MSG_OOB option if you want to receive out of band data. Refer to the "Sending and Receiving Out of Band Data" section of chapter 3, "Advanced Topics for Stream Sockets," for more information.

Use the MSG_PEEK option to preview incoming data. If this option is set on a `recv`, any data returned remains in the socket buffer as though it had not been read yet. The next `recv` returns the same data.

When to Receive Data

The server or client process should receive data after connection is established. Refer to the `recv(2)` man page for more information on `recv`.

Closing a Socket

In most applications, you do not have to worry about cleaning up your sockets. When you exit your program and your process terminates, the sockets are closed for you. If you need to close a socket while your program is still running, use the `close` system call.

For example, you may have a daemon process that uses `fork` to create the server process. The daemon process creates the BSD sockets connection and then passes the socket descriptor to the server. You then have more than one process with the same socket descriptor. The daemon process should do a `close` of the socket descriptor to avoid keeping the socket open once the server is through with it. Because the server performs the work, the daemon does not use the socket after the `fork`.

`close` decrements the file descriptor reference count. Once this occurs, the calling process can no longer use that file descriptor.

When the last `close` is executed on a socket descriptor, any unsent data are sent before the socket is closed. Any unreceived data are lost. This delay in closing the socket can be controlled by the socket option SO_LINGER. Additional options for closing sockets are discussed in chapter 3, "Advanced Topics for Stream Sockets."

For syntax and details on `close`, refer to the `close(2)` man page.

Example Using Internet Stream Sockets

NOTE These programs are provided as examples only of stream socket usage and are not Hewlett-Packard supported products.

These program examples demonstrate how to set up and use internet stream sockets. These sample programs are in the `/usr/lib/demos/networking/socket` directory. The client program is intended to run in conjunction with the server program. The client program requests a service called `example` from the server program.

The server process receives requests from the remote client process, handles the request and returns the results to the client process. Note that the server:

- Uses the wildcard address for the listen socket.
- Uses the `ntohs` address conversion call to show porting to a host that requires it.
- Uses the SO_LINGER option for a graceful disconnect.

The client process creates a connection, sends requests to the server process and receives the results from the server process. Note that the client:

- Uses `shutdown` to indicate that it is done sending requests.
- Uses `getsockname` to see what socket address was assigned to the local socket by the host.
- Uses the `ntohs` address conversion call to show porting to a host that requires it.

Before you run the example programs, make the following entry in the two host's `/etc/services` files:

```
example  22375/tcp
```

The source code for these two programs follows

```
/*
 *                      S E R V . T C P
 *
 * This is an example program that demonstrates the use of stream
 * sockets as a BSD Sockets mechanism.  This contains the server,
 * and is intended to operate in conjunction with the client
 * program found in client.tcp.  Together, these two programs
```

```
* demonstrate many of the features of sockets, as well as good
* conventions for using these features.
*
* This program provides a service called "example". In order for
* it to function, an entry for it needs to exist in the
* /etc/services file.  The port address for this service can be
* any port number that is likely to be unused, such as 22375,
* for example.  The host on which the client will be running
* must also have the same entry (same port number) in its
* /etc/services file.
*
*/

#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <signal.h>
#include <stdio.h>
#include <netdb.h>

int s;                  /* connected socket descriptor */
int ls;                 /* listen socket descriptor */

struct hostent *hp;     /* pointer to host info for remote host */
struct servent *sp;     /* pointer to service information */

long timevar;           /* contains time returned by time() */
char *ctime();          /* declare time formatting routine */

struct linger linger = {1,1};
                         /* allow lingering, graceful close; */
                         /* used when setting SO_LINGER */

struct sockaddr_in myaddr_in;  /* for local socket address */
struct sockaddr_in peeraddr_in;/* for peer socket address */

/*

*                            M A I N
*
* This routine starts the server.  It forks, leaving the child
* to do all the work, so it does not have to be run in the
* background.It sets up the listen socket, and for each incoming
* connection, it forks a child process to process the data.
* It will loop forever, until killed by a signal.
*/
main(argc, argv)
int argc;
char *argv[];
{
     int addrlen;
        /* clear out address structures */
     memset ((char *)&myaddr_in, 0, sizeof(struct sockaddr_in));
     memset ((char *)&peeraddr_in, 0, sizeof(struct sockaddr_in));
             /* Set up address structure for the listen socket. */
     myaddr_in.sin_family = AF_INET;
     /* The server should listen on the wildcard address,
      * rather than its own internet address.  This is
      * generally good practice for servers, because on
      * systems which are connected to more than one
```

```
     * network at once will be able to have one server
      * listening on all networks at once.  Even when the
      * host is connected to only one network, this is good
      * practice, because it makes the server program more
      * portable.
      */
myaddr_in.sin_addr.s_addr = INADDR_ANY;
      /* Find the information for the "example" server
      * in order to get the needed port number.
      */
sp = getservbyname ("example", "tcp");
if (sp == NULL) {
      fprintf(stderr, "%s: host not found ",
                     argv[0]);

      exit(1);
}
myaddr_in.sin_port = sp->s_port;

      /* Create the listen socket. */
ls = socket (AF_INET, SOCK_STREAM, 0);
if (ls == -1) {
      perror(argv[0]);
      fprintf(stderr, "%s: unable to create socket\n" , argv[0]);
      exit(1);
}
      /* Bind the listen address to the socket. */
if (bind(ls, &myaddr_in, sizeof(struct sockaddr_in)) == -1) {
      perror(argv[0]);
      fprintf(stderr, "%s: unable to bind address\n", argv[0]);
      exit(1);
}
      /* Initiate the listen on the socket so remote users
      * can connect.  The listen backlog is set to 5. 20
      */
if (listen(ls, 5) == -1) {
      perror(argv[0]);
      fprintf(stderr, "%s: unable to listen on socket\n",argv[0]);
      exit(1);
}

      /* Now, all the initialization of the server is
      * complete, and any user errors will have already
      * been detected.  Now we can fork the daemon and
      * return to the user.  We need to do a setpgrp
      * so that the daemon will no longer be associated
      * with the user's control terminal.  This is done
      * before the fork, so that the child will not be
      * a process group leader.  Otherwise, if the child
      * were to open a terminal, it would become associated
      * with that terminal as its control terminal.  It is
      * always best for the parent to do the setpgrp.
      */
setpgrp();

switch (fork()) {
case -1:
      /* Unable to fork, for some reason. */
      perror(argv[0]);
```

```
        fprintf(stderr, "%s: unable to fork daemon\n", argv[0]);
        exit(1);

case 0:
        /* The child process (daemon) comes here. */
        /* Close stdin and stderr so that they will not
        * be kept open.  Stdout is assumed to have been
        * redirected to some logging file, or /dev/null.
        * From now on, the daemon will not report any
        * error messages.  This daemon will loop forever,
        * waiting for connections and forking a child
        * server to handle each one.
        */
       fclose(stdin);
       fclose(stderr);
              /* Set SIGCLD to SIG_IGN, in order to prevent
               * the accumulation of zombies as each child
               * terminates.  This means the daemon does not
               * have to make wait calls to clean them up.
               */
       signal(SIGCLD, SIG_IGN);
       for(;;) {
               /* Note that addrlen is passed as a pointer
                * so that the accept call can return the
                * size of the returned address.
                */
               addrlen = sizeof(struct sockaddr_in);
                      /* This call will block until a new
                      * connection arrives. Then, it will
                      * return the address of the connecting
                      * peer, and a new socket descriptor,
                      * s, for that connection.
                      */
               s = accept(ls, &peeraddr_in, &addrlen);
               if ( s == -1) exit(1);
               switch (fork()) {
               case -1:        /* Can't fork, just continue. */
                      exit(1);
               case 0:         /* Child process comes here. */
                      server();
                      exit(0);
               default:        /* Daemon process comes here. */
                               /* The daemon needs to close the
                                * the new accept socket after
                                * forking the child. This
                                * prevents daemon from running
                                * out of file descriptors.
                                * It also means that when the
                                * server closes the socket,
                                * that it will allow socket
                                * to be destroyed since it
                                * will be the last close.
                                */
                      close(s);
               }
       }
default:              /* Parent process comes here. */
       exit(0);
}
```

```
}
 /*
 *                              S E R V E R
 *
 *      This is the actual server routine that the daemon forks
 *      to handle each individual connection.  Its purpose is
 *      to receive the request packets from the remote client,
 *      process them, and return the results to the client.
 *      It will also write some logging information to stdout.
 *
 */
server()
{
        int reqcnt = 0;    /* keeps count of number of requests */
        char buf[10];      /* This example uses 10 byte messages. */
        char *inet_ntoa();
        char *hostname; /* points to remote host's name string */
        int len, len1;
           /* Close the listen socket inherited from the daemon. */
        close(ls);

                /* Look up the host information for the remote host
                 * we have connected with. Its internet address
                 * was returned by the accept call, in the main
                 * daemon loop above.
                 */
        hp = gethostbyaddr ((char *) &peeraddr_in.sin_addr,
                                sizeof (struct in_addr),
                                peeraddr_in.sin_family);

        if (hp == NULL) {
                   /* The info is unavailable for the remote host.
                    * Just format its internet address to be
                    * printed in the logging information.  The
                    * address will be shown in internet dot format.
                    */
                hostname = inet_ntoa(peeraddr_in.sin_addr);
        } else {
                hostname = hp->h_name;  /* point to host's name */
        }
                /* Log a startup message. */
        time (&timevar);
                /* The port number must be converted first to
                 * host byte order before printing. On most hosts,
                 * this is not necessary, but the ntohs() call is
                 * included here so that this program could easily
                 * be ported to a host that does require it.
                 */
        printf("Startup from %s port %u at %s",
          hostname, ntohs(peeraddr_in.sin_port), ctime(&timevar));

                /* Set the socket for a lingering, graceful close.
                 * Since linger was set to 1 above, this will
                 * cause a final close of this socket to wait
                 * until all of the data sent on it has been
                 * received by the remote host.
                */
        if (setsockopt(s, SOL_SOCKET, SO_LINGER, &linger,
                                      sizeof(linger)) == -1) {
errout:         printf("Connectionwith%saborterror\n",hostname);
```

```
        exit(1);
}

      /* Go into a loop, receiving requests from the
       * remote client.  After the client has sent the
       * last request, it will do a shutdown for sending,
       * which causes an end-of-file condition to appear
       * on this end of the connection.  After all of the
       * client's requests have been received, the next
       * recv call will return zero bytes, signalling an
       * end-of-file condition. This is is how the server
       * will know that no more requests will follow
       * and the loop will be exited.*/
while (len = recv(s, buf, 10, 0)) {
       if (len == -1) goto errout; /* error from recv */
             /* The reason this while loop exists is that
              * there is a remote possibility of the above
              * recv returning less than 10 bytes. This is
              * because a recv returns as soon as there is
              * some data, and will not wait for all of the
              * requested data to arrive. Since 10 bytes is
              * relatively small compared to the allowed TCP
              * packet sizes, a partial receive is unlikely.
              * If this example had used 2048 bytes requests
              * instead, a partial receive would be far more
              * likely. This loop will keep receiving until
              * all 10 bytes have been received, thus
              * guaranteeing that the next recv at the top
              * of the loop will start at the beginning
              * of the next request.
              */
        while (len < 10) {
                len1 = recv(s, &buf[len], 10-len, 0);
                if (len1 == -1) goto errout;
                len += len1;
        }
                /* Increment the request count. */
        reqcnt++;
                /* This sleep simulates the processing of
                 * the request that a real server might do.
                 */
        sleep(1);
                /* Send a response back to the client. */
        if (send(s, buf, 10, 0) != 10) goto errout;
}

        /* The loop has terminated, because there are no
         * more requests to be serviced. As above, this
         * close will block until all of the sent replies
         * have been received by the remote host. Lingering
         * on the close is so the server will have a better
         * idea when the remote has picked up all the data.
         * This allows the start and finish times printed
         * in the log file to more accurately reflect
         * the length of time this connection was used.
         */
close(s);

        /* Log a finishing message. */
time (&timevar);
```

```
                /* The port number must be converted first to
                 * host byte order before printing.  On most hosts,
                 * this is not necessary, but the ntohs() call is
                 * included here so this program could easily
                 * be ported to a host that does require it.
                 */
        printf("Completed %s port %u, %d requests, at %s\n",
               hostname, ntohs(peeraddr_in.sin_port), reqcnt,
               ctime(&timevar));
}

/*
 *                      C L I E N T . T C P
 *
 *      This example program demonstrates the use of stream
 *      sockets as a BSD Sockets mechanism.  This contains the client,
 *      and is intended to operate in conjunction with the server
 *      program found in serv.tcp.  Together, these two programs
 *      demonstrate many of the features of sockets, as well as
 *      good conventions for using these features.
 *
 *      This program requests a service called "example".  For it
 *      to function, an entry needs to exist in the /etc/services
 *      file.  The port address for this service can be any port
 *      number that is not used, such as 22375, for example. The
 *      host on which the server will be running must also have the
 *      same entry (same port number) in its /etc/services file.
 *
 */

#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <stdio.h>
#include <netdb.h>

int s;                          /* connected socket descriptor */

struct hostent *hp;             /* pointer to host info for remote host */
struct servent *sp;             /* pointer to service information */

long timevar;                   /* contains time returned by time() */
char *ctime();                  /* declare time formatting routine */

struct sockaddr_in myaddr_in;   /* for local socket address */
struct sockaddr_in peeraddr_in; /* for peer socket address */

/*
 *                      M A I N
 *
 *      This routine is the client that requests service from the
 *      remote example server. It creates a connection, sends a few
 *      of requests, shuts down the connection in one direction to
 *      signal the server about the end of data, and then receives
 *      all of the responses. Status will be written to stdout.
 *
 *         The name of the system to which the requests will be sent
 *         is given as a parameter to the command.
 */
main(argc, argv)
```

```
int argc;
char *argv[];
{
        int addrlen, i, j;

                /* This example uses 10 byte messages. */
        char buf[10];

    if (argc != 2) {
        fprintf(stderr, "Usage:%s <remote host>\n" argv[0];
        exit(1);
    }
        /* clear out address structures */
    memset ((char *)&myaddr_in, 0, sizeof(struct sockaddr_in));
    memset ((char *)&peeraddr_in, 0, sizeof(struct sockaddr_in));

        /* Set up the peer address to which we will connect. */
    peeraddr_in.sin_family = AF_INET;
        /* Get the host information for the hostname that the
         * user passed in.
         */
    hp = gethostbyname (argv[1]);
         /* argv[1] is the host name. */
    if (hp == NULL) {
           fprintf(stderr, "%s: %s not found in /etc/hosts\n",
                            argv[0], argv[1]);
           exit(1);
    }
    peeraddr_in.sin_addr.s_addr=
      ((struct in_addr*)(hp->h_addr)->s_addr;
              /* Find the information for the "example" server
               * in order to get the needed port number.
               */
    sp = getservbyname ("example", "tcp");
    if (sp == NULL) {
            fprintf(stderr, "%s: example not found in
            /etc/services\n", argv[0]);
            exit(1);
    }
    peeraddr_in.sin_port = sp->s_port;

            /* Create the socket. */
    s = socket (AF_INET, SOCK_STREAM, 0);
    if (s == -1) {
         perror(argv[0]);
         fprintf(stderr,"%s: unable to create socket\n", argv[0])
         exit(1);
    }
           /* Try to connect to the remote server at the
            * address which was just built into peeraddr.
            */
  if (connect(s, &peeraddr_in, sizeof(struct sockaddr_in)) ==-1) {
           perror(argv[0]);
           fprintf(stderr, "%s: unable to connect to remote\n",
                                argv[0]);
           exit(1);
}
                /* Since the connect call assigns a random address
                 * to the local end of this connection, let's use
                 * getsockname to see what it assigned.  Note that
```

```
                 * addrlen needs to be passed in as a pointer,
                 * because getsockname returns the actual length
                 * of the address.
                 */
addrlen = sizeof(struct sockaddr_in);
if (getsockname(s, &myaddr_in, &addrlen) == -1) {
       perror(argv[0]);
       fprintf(stderr, "%s: unable to read socket address\n",
                             argv[0]);
       exit(1);
}

                /* Print out a startup message for the user. */
        time(&timevar);
                /* The port number must be converted first to
                 * host byte order before printing.  On most hosts,
                 * this is not necessary, but the ntohs() call is
                 * included here so this program could easily be
                 * ported to a host that does require it.
                 */
        printf("Connected to %s on port %u at %s",
                        argv[1], ntohs(myaddr_in.sin_port),
ctime(&timevar));

                /* This sleep simulates any preliminary processing
                 * that a real client might do here.
                 */
        sleep(5);

                /* Send out all the requests to the remote server.
                 * In this case five are sent but any random number
                 * could be used. The first four bytes of buf are
                 * set up to contain the request number. This
                 * number will be returned in the server's reply.
                 */

                /* CAUTION: If you increase the number of requests
                 * sent or the size of the requests, you should be
                 * aware that you could encounter a deadlock
                 * situation.  Both the client's and server's
                 * sockets can only queue a limited amount of
                 * data on their receive queues.
                 */
        for (i=1; i<=5; i++) {
              *buf = i;
              if (send(s, buf, 10, 0) != 10) {
                 fprintf(stderr, "%s: Connection aborted on error ",
                                       argv[0]);
                        fprintf(stderr, "on send number %d\n", i);
                        exit(1);
                }
        }

                /* Now, shutdown the connection for further sends.
                 * This causes the server to receive an end-of-file
                 * condition after receiving all the requests that
                 * have just been sent, indicating that no further
                 * requests will be sent.
                 */
```

```
        if (shutdown(s, 1) == -1) {
                perror(argv[0]);
        fprintf(stderr, "%s: unable to shutdown socket\n",argv[0]);
                exit(1);
        }

     /* Start receiving all the replys from the server.
      * This loop will terminate when the recv returns
      * zero, which is an end-of-file condition.  This
      * will happen after the server has sent all of its
      * replies, and closed its end of the connection.
      */
      while (i = recv(s, buf, 10, 0)) {
                if (i == -1) {
errout:                 perror(argv[0]);
        fprintf(stderr, "%s: error reading result\n",argv[0]);
                      exit(1);
      }

      /* The reason this while loop exists is that there
       * is a remote possibility of the above recv returning
       * less than 10 bytes.  This is because a recv returns
       * as soon as there is some data, and will not wait for
       * all of the requested data to arrive.  Since 10 bytes
       * is relatively small compared to the allowed TCP
       * packet sizes, a partial receive is unlikely.  If
       * this example had used 2048 bytes requests instead,
       * a partial receive would be far more likely.
       * This loop will keep receiving until all 10 bytes
       * have been received, thus guaranteeing that the
       * next recv at the top of the loop will
       * start at the begining of the next reply.
       */
      while (i < 10) {
               j = recv(s, &buf[i], 10-i, 0);
               if (j == -1) goto errout;
               i += j;
      }
               /* Print out message indicating the
               * identity of this reply.
               */
      printf("Received result number %d\n", *(int *)buf);
  }

      /* Print message indicating completion of task. */

    time(&timevar);
    printf("All done at %s", ctime(&timevar));
}
```

3 Advanced Topics for Stream Sockets

Advanced Topics for Stream Sockets

This chapter explains the following:

- Socket options.
- Synchronous I/O multiplexing with `select`.
- Sending and receiving data asynchronously.
- Nonblocking I/O.
- Using shutdown.
- Using `read` and `write` to make stream sockets transparent.
- Sending and receiving out of band data.

Socket Options

The operation of sockets is controlled by socket level options. The following options are supported for internet stream sockets:

- SO_REUSEADDR
- SO_KEEPALIVE
- SO_DONTROUTE
- SO_SNDBUF
- SO_RCVBUF
- SO_LINGER
- SO_USELOOPBACK
- SO_OOBINLINE
- SO_SNDLOWAT
- SO_RCVLOWAT
- SO_SNDTIMEO
- SO_RCVTIMEO
- SO_TYPE
- SO_ERROR
- SO_BROADCAST
- SO_REUSEPORT

All of these options may be used on either AF_INET or AF_UNIX sockets; the following, however, are really INET specific in their function and will not change UNIX socket behavior.

- SO_KEEPALIVE
- SO_REUSEADDR
- SO_DONTROUTE
- SO_USELOOPBACK
- SO_OOBINLINE

- SO_SNDTIMEO
- SO_RCVTIMEO
- SO_BROADCAST
- SO_REUSEPORT

In addition, the SO_DEBUG option is supported for compatibility only; it has no functionality.

Options for protocol levels are described in the individual protocol manual pages, such as `tcp(7p)`, `udp(7p)`, and `ip(7p)`.

The next section describes how to get the current value of a socket option and to set socket options, followed by a description of each available option. Refer to chapter 6 for a description of the SO_BROADCAST option.

Getting and Setting Socket Options

The socket options are defined in the `sys/socket.h` file. You can get the current status of an option with the `getsockopt` call, and you can set the value of an option with the `setsockopt` call.

`getsockopt` and its parameters are described in the following table:

Include files:

```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:

```
getsockopt(s, level, optname, optval, optlen)
int s, level, optname;
char *optval;
int *optlen;
```

Parameter	Contents	INPUT Value	OUTPUT Value
s	socket descriptor	socket descriptor for which option values are to be returned	unchanged
level	protocol level	SOL_SOCKET	unchanged
optname	name of option	supported option name	unchanged
optval	pointer to current value of option	pointer to buffer where option's current value is to be returned	pointer to buffer that contains current option value
optlen	pointer to length of optval	pointer to maximum number of bytes to be returned by optval	pointer to actual size of optval returned

Function result: 0 if the option is set. If getsockopt fails for any reason, the function returns -1, and the option is not returned. An error code is stored in errno.

Example:

```
len = sizeof (optval))
getsockopt(s, SOL_SOCKET, SO_REUSEADDR, &optval, &len;)
```

`optval` may never be zero. It must always point to data sent with the socket option and must always be at least the size of an integer.

The following socket options set socket parameter values. *optval* is an integer containing the new value:

- SO_SNDBUF
- SO_RCVBUF
- SO_SNDLOWAT
- SO_RCVLOWAT
- SO_SNDTIMEO

- SO_RCVTIMEO

The following socket options toggle socket behavior. optval is an integer containing a boolean flag for the behavior (1 = on, 0 = off):

- SO_KEEPALIVE
- SO_DEBUG
- SO_DONTROUTE
- SO_USELOOPBACK
- SO_REUSEADDR
- SO_OOBINLINE
- SO_REUSEPORT

The SO_LINGER option is a combination. It sets a linger value, and also toggles linger behavior on and off. In previous releases SO_DONTLINGER was supported. For SO_LINGER, optval points to a struct linger, defined in /usr/include/sys/socket.h. The structure contains an integer boolean flag to toggle behavior on/off, and an integer linger value. Refer to the getsockopt(2) man page for more information on getsockopt.

setsockopt and its parameters are described in the following table:

Include files:

```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:

```
setsockopt(s, level, optname, optval, optlen)
int s, level, optname;
char *optval;
int optlen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor	socket descriptor for which options are to be set
level	protocol level	SOL_SOCKET

Parameter	Description of Contents	INPUT Value
optname	name of option	supported option name
optval	pointer to option input value	Must be at least size of (int). Holds either value to be set or boolean flag
optlen	length of optval	size of optval

Function result: 0 if setsockopt is successful, −1 if failure occurs.

Example: See the description of the SO_REUSEADDR option for an example.

Refer to the `setsockopt(2)` man page for more information on `setsockopt`.

SO_REUSEADDR

This option is AF_INET socket-specific.

SO_REUSEADDR enables you to restart a daemon which was killed or terminated.

This option modifies the rules used by bind to validate local addresses, but it does not violate the uniqueness requirements of an association. SO_REUSEADDR modifies the bind rules only when a wildcard Internet Protocol (IP) address is used in combination with a particular protocol port. The host still checks at connection time to be sure any other sockets with the same local address and local port do not have the same remote address and remote port. Connect fails if the uniqueness requirement is violated.

Example of the SO_REUSEADDR Option

A network daemon server is listening on a specific port: port 2000. If you executed `netstat an`, part of the output would resemble:

```
Active connections (including   servers)
Proto Recv-Q Send-Q  Local Address Foreign Address (state)
tcp        0      0    *.2000       *.*             LISTEN
```

Network Daemon Server Listening at Port 2000. When the network daemon accepts a connection request, the accepted socket will bind to port 2000 and to the address where the daemon is running (e.g. `192.6.250.100`). If you executed `netstat an`, the output would resemble:

```
Active connections (including servers)
Proto Recv-Q Send-Q  Local Address Foreign Address (state)
tcp 0 0 192.6.250.100.2000 192.6.250.101.4000 ESTABLISHED
tcp 0 0 *.2000 *.* LISTEN
```

New Connection Established, Daemon Server Still Listening. Here the network daemon has established a connection to the client (`192.6.250.101.4000`) with a new server socket. The original network daemon server continues to listen for more connection requests.

If the listening network daemon process is killed, attempts to restart the daemon fail if SO_REUSEADDR is not set. The restart fails because the daemon attempts to bind to port 2000 and a wildcard IP address (e.g. `*.2000`). The wildcard address matches the address of the established connection (`192.6.250.100`), so the `bind` aborts to avoid duplicate socket naming.

When SO_REUSEADDR is set, `bind` ignores the wildcard match, so the network daemon can be restarted. An example usage of this option is:

```
int optval = 1;
setsockopt (s, SOL_SOCKET, SO_REUSEADDR, &optval,
sizeof(optval));
bind (s, &sin, sizeof(sin));
```

SO_KEEPALIVE

This option is AF_INET socket-specific.

This option enables the periodic transmission of messages on a connected socket. This occurs at the transport level and does not require any work in your application programs.

If the peer socket does not respond to these messages, the connection is considered broken. The next time one of your processes attempts to use a connection that is considered broken, the process is notified (with a SIGPIPE signal if you are trying to send, or an end-of-file condition if you are trying to receive) that the connection is broken.

SO_DONTROUTE

This option is AF_INET socket-specific.

SO_DONTROUTE indicates that outgoing messages should bypass the standard routing facilities. Instead, messages are directed to the appropriate network interface according to the network portion of the destination address.

SO_SNDBUF

SO_SNDBUF changes the send socket buffer size. Increasing the send socket buffer size allows a user to send more data before the user's application will block, waiting for more buffer space.

NOTE Increasing buffer size to send larger portions of data before the application blocks *may* increase throughput, but the best method of tuning performance is to experiment with various buffer sizes.

You can increase a stream socket's buffer size at any time, but decrease it only prior to establishing a connection. The maximum buffer size for stream sockets is 262144 bytes. Here is an example:

```
int result;
int buffsize = 10,000;
result = setsockopt(s, SOL_SOCKET, SO_SNDBUF, &buffsize,
sizeof(buffsize));
```

SO_RCVBUF

SO_RCVBUF changes the receive socket buffer size.

You can increase a stream socket's buffer size at any time, but decrease it only prior to establishing a connection. The maximum buffer size for stream sockets is 262144 bytes. Here is an example:

```
int result;
int buffsize = 10,000;
result=setsockopt(s,SOL_SOCKET,SO_RCVBUF,&buffsize,sizeof
(buffsize));
```

Table 3-1 **Summary Information for Changing Socket Buffer Size**

Socket Type (Protocol)	stream (TCP)
When Buffer Size Increase Allowed	at any time
When Buffer Size Decrease Allowed	only prior to establishing a connection
Maximum Buffer Size	262144 bytes

SO_LINGER

SO_LINGER controls the actions taken when a `close` is executed on a socket that has unsent data. This option can be cleared by toggling. The default is off.

The linger timeout interval is set with a parameter in the `setsockopt` call. The only useful values are zero and nonzero:

- If l_onoff is zero, `close` returns immediately, but any unsent data is transmitted (after `close` returns).
- If l_onoff is nonzero and l_linger is zero, `close` returns immediately, and any unsent data is discarded.
- If l_onoff is nonzero and l_linger is nonzero, `close` does not return until all unsent data is transmitted (or the connection is closed by the remote system).

In the default case (SO_LINGER is off), `close` is not blocked. The socket itself, however, goes through graceful disconnect, and no data is lost. Here is an example:

```
int result;
struct linger linger;
linger.l_onoff = 1;
       /*0 = off (l_linger ignored), nonzero = on */
linger.l_linger =1;
       /*0 = discard data, nonzero = wait for data sent */
result = setsockopt(s, SOL_SOCKET, SO_LINGER, &linger,
sizeof(linger));
```

Table 3-2 Summary of Linger Options on Close

Socket Option	Option Set	Linger Interval	Graceful Close	Hard Close	Wait for Close	Does Not Wait for Close
SO_LINGER	off	don't care	x			x
SO_LINGER	on	zero		x		x
SO_LINGER	on	nonzero	x		x	

SO_USELOOPBACK

This option is not applicable to UNIX Domain sockets.

SO_USELOOPBACK directs the network layer (IP) of networking code to use the local loopback address when sending data from this socket. Use this option only when all data sent will also be received locally.

SO_OOBINLINE

This option is not applicable to UNIX Domain sockets.

This option enables receipt of out-of-band data inline. Normally, OOB data is extracted from the data stream and must be read with the MSG_OOB flag specified in the `recv()` call. When SO_OOBINLINE is specified, OOB data arriving at that socket remains inline and can be read without MSG_OOB specified.

In both cases, a normal `read()` or `recv()` which would read past the OOB mark will halt at the mark, instead leaving the OOB byte the next byte to be read.

SO_SNDLOWAT

This option allows the user to set or fetch the low water mark for the socket's send socket buffer. At present, this option is not used. It is supported in anticipation of future use.

SO_RCVLOWAT

This option allows the user to set or fetch the low water mark for the socket's receive socket buffer. At present, this option is not used. It is supported in anticipation of future use.

SO_SNDTIMEO

This option allows the user to set or fetch the timeout value for a socket's send socket buffer.At present, this option is not used. It is supported in anticipation of future use.

SO_RCVTIMEO

This option allows the user to set or fetch the timeout value for the socket's receive socket buffer. At present, this option is not used. It is supported in anticipation of future use.

SO_TYPE

This option is used to return the socket type (e.g., stream, datagram, etc.). Use this option only with the `getsockopt` system call.

SO_ERROR

This option is used to get and clear any error that has occurred on the socket. Use this option only with the `getsockopt` system call.

SO_BROADCAST

This option is not supported for UNIX Domain sockets. Setting this option allows the user to send datagrams to a broadcast address. A broadcast address is defined as an internet address whose local address portion is all 1s.

SO_REUSEPORT

This option is AF_INET socket-specific. This option allows multiple processes to share a port. All incoming multicast or broadcast UDP datagrams that are destined for the port are delivered to all sockets that

are bound to the port. All processes that share the port must specify this option. For more information on using this option, see "Sending and Receiving IP Multicast Datagrams," in Chapter 5.

Synchronous I/O Multiplexing with Select

The `select` system call can be used with sockets to provide a synchronous multiplexing mechanism. The system call has several parameters which govern its behavior. If you specify a zero pointer for the timeout parameter, `select` will block until one or more of the specified socket descriptors is ready. If timeout is a non-zero pointer, it specifies a maximum interval to wait for the selection to complete.

A `select` of a socket descriptor for **reading** is useful on:

- A connected socket, because it determines when data have arrived and are ready to be read without blocking; use the FIONREAD parameter to the `ioctl` system call to determine exactly how much data are available.
- A listening socket, because it determines when you can accept a connection without blocking.
- Any socket to detect if an error has occurred on the socket.

A `select` of a socket descriptor for **writing** is useful on:

- A connecting socket, because it determines when a connection is complete.
- A connected socket, because it determines when more data can be sent without blocking. This implies that at least one byte can be sent; there is no way, however, to determine exactly how many bytes can be sent.
- Any socket to detect if an error has occurred on the socket.

`select` for exceptional conditions will return true for BSD sockets if out-of-band data is available to be read. `select` will always return true for sockets which are no longer capable of being used (e.g. if a `close` or `shutdown` system call has been executed against them).

`select` is used in the same way as in other applications. Refer to the `select(2)` man page for information on how to use `select`.

The following example illustrates the `select` system call. Since it is possible for a process to have more than 32 open file descriptors, the bit masks used by `select` are interpreted as arrays of integers. The header file `sys/types.h` contains some useful macros to be used with the `select()` system call, some of which are reproduced below.

```
/*
* These macros are used with select(). select() uses bit masks of
* file descriptors in long integers.  These macros manipulate such
* bit fields (the file system macros use chars).  FD_SETSIZE may
* be defined by the user, but must be = u.u_highestfd + 1. Since
* the absolute limit on the number of per process open files is
* 2048, FD_SETSIZE must be large enough to accommodate this many
* file descriptors.
* Unless the user has this many files opened, FD_SETSIZE should
* be redefined to a smaller number.
*/

typedef long fd_mask
#define NFDBITS (sizeof (fd_mask) * 8   /* 8 bits per byte
#define howmany (x,y)  (((x)+((y)-1))/(y))
typedef struct fd_set  {
   fd_mask fds_bits [howmany (FD_SETSIZE, NFDBITS)]; */
}  fd_set;
#define FD_SET(n,p)  ((p)->fds_bits[(n)/NFDBITS]
#|= (1 << ((n) % NFDBITS)))
#define FD_CLR(n,p)   ((p)->fds_bits[(n)/NFDBITS]
#&= ~(1 << ((n) % NFDBITS)))
#define FD_ISSET(n,p)  ((p) ->fds_bits[(n)/NFDBITS]
#& (1 << ((n) % NFDBITS)))
#define FD_ZERO(p)  memset((char *) (p), (char) 0, sizeof (*(p)))

do_select(s)
int s;                      /* socket to select on, initialized */
{
   struct fd_set read_mask, write_mask; /* bit masks */
   int nfds;                    /* number to select on */
   int nfd;                     /* number found */

   for (;;) {                   /* for example... */
       FD_ZERO(&read_mask);     /* select will overwrite on return */
       FD_ZERO(&write_mask);
       FD_SET(s, &read_mask);   /* we care only about the socket */
       FD_SET(s, &write_mask);
       nfds = s+1;              /* select descriptors 0 through s */
       nfd = select(nfds, &read_mask, &write_mask, (int *) 0,
               (struct timeval *) 0);/* will block */
       if (nfd == -1) {
              perror( "select: unexpected condition" );
```

```
                exit(1);
        }
        if (FD_ISSET(s, &read_mask))
                do_read(s);         /* something to read on socket s */
                                    /* fall through as maybe more to do */

        if (FD_ISSET(s, &write_mask))
                do_write(s); /* space to write on socket s */
        }
}
```

Sending and Receiving Data Asynchronously

Asynchronous sockets allow a user program to receive an SIGIO signal when the state of the socket changes. This state change can occur, for example, when new data arrives. Currently the user would have to issue a `select` system call in order to determine if data were available. If other processing is required of the user program, the need to call `select` can complicate an application by forcing the user to implement some form of polling, whereby all sockets are checked periodically. Asynchronous sockets allow the user to separate socket processing from other processing, eliminating polling altogether. `select` may still be required to determine exactly why the signal is being delivered, or to which socket the signal applies.

Generation of the SIGIO signal is protocol dependent. It mimics the semantics of `select` in the sense that the signal is generated whenever `select` returns true. It is generally accepted that connectionless protocols deliver the signal whenever a new packet arrives. For connection oriented protocols, the signal is also delivered when connections are established or broken, as well as when additional outgoing buffer space becomes available. Be warned that these assertions are guidelines only; any signal handler should be robust enough to handle signals in unexpected situations.

The delivery of the SIGIO signal is dependent upon two things. First, the socket state must be set as asynchronous; this is done using the FIOASYNC flag of the `ioctl` system call. Second, the process group (pgrp) associated with the socket must be set; this is done using the SIOCSPGRP flag of `ioctl`. The sign value of the pgrp can lead to various signals being delivered. Specifically, if the pgrp is positive, this implies that a signal should be delivered to the process whose PID is the absolute value of the pgrp. If the pgrp is negative, a signal should be delivered to the process group identified by the absolute value of the pgrp.

Any application that chooses to use asynchronous sockets must explicitly activate the described mechanism. The SIGIO signal is a "safe" signal in the sense that if a process is unprepared to handle it, the default action is to ignore it. Thus any existing applications are immune to spurious signal delivery.

Notification that out-of-band data has been received is also done asynchronously; see the section "Sending and Receiving Out-of-band Data" in this chapter for more details.

The following example sets up an asynchronous SOCK_STREAM listen socket. This is typical of an application that needs to be notified when connection requests arrive.

```
int ls;           /* SOCK_STREAM listen socket initialized */
int flag = 1;     /* for ioctl, to turn on async */
int iohndlr();    /* the function which handles the SIGIO */

signal (SIGIO, iohndlr); /* set up the handler */

if (ioctl (ls, FIOASYNC, &flag) == -1) {
       perror ("can't set async on socket");
       exit(1);
}
flag = -getpid();/* process group negative = deliver to process */
if (ioctl (ls, SIOCSPGRP, &flag) == -1) {
       perror ("can't set pgrp");
       exit(1);
}
       /* signal can come any time now */
```

The following example illustrates the use of process group notification. Note that the real utility of this feature is to allow multiple processes to receive the signal, which is not illustrated here. For example, the socket could be of type SOCK_DGRAM; a signal here can be interpreted as the arrival of a service-request packet. Multiple identical servers could be set up, and the first available one could receive and process the packet.

```
int flag = 1;                 /* ioctl to turn on async */
int iohndlr();
signal (SIGIO, iohndlr);

setpgrp();                    /* set my processes' process group */
if (ioctl (s, FIOASYNC, &flag) == -1) {
       perror ("can't set async on socket");
       exit(1);
}
flag = getpid(); /* process group + = deliver to each process in group */
if (ioctl (s, SIOCSPGRP, &flag) == -1) {
       perror ("can't set pgrp");
       exit(1);
}
       /* signal can come any time now */
```

Nonblocking I/O

Sockets are created in blocking mode I/O by default. You can specify that a socket be put in nonblocking mode by using the `ioctl` system call with the FIOSNBIO request. Here is an example:

```
#include <sys/ioctl.h>
...
ioctl(s, FIOSNBIO, &arg);
```

`arg` is a pointer to `int`:

- When `int` equals 0, the socket is changed to blocking mode.
- When `int` equals 1, the socket is changed to nonblocking mode.

If a socket is in nonblocking mode, the following calls are affected:

`accept`	If no connection requests are present, `accept` returns immediately with the EWOULDBLOCK error.
`connect`	If the connection cannot be completed immediately, `connect` returns with the EINPROGRESS error.
`recv`	If no data are available to be received, `recv` returns the value –1 and the EWOULDBLOCK error. This is also true for `read`.
`send`	If there is no available buffer space for the data to be transmitted, `send` returns the value -1 and the EWOULDBLOCK error. This is also true for `write`.

The O_NDELAY flag for `fcntl(2)` is also supported. If you use this flag and there are no data available to be received on a `recv`, `recvfrom`, `recvmsg`, or `read` call, the call returns immediately with the value of 0. If you use the O_NONBLOCK flag, the call returns immediately with the value of -1 and the EAGAIN error. This is the same as returning an end-of-file condition. This is also true for `send`, `sendto`, `sendmsg`, and `write` if there is not enough buffer space to complete the send.

NOTE The O_NDELAY and O_NONBLOCK flags have precedence over the FIOSNBIO flag. Setting both the O_NDELAY and O_NONBLOCK flags is not allowed.

Using Shutdown

When your program is done reading or writing on a particular socket connection, you can use `shutdown` to bring down a part of the connection.

When one process uses `shutdown` on a socket descriptor, all other processes with the same socket descriptor are affected. `shutdown` causes all or part of a full-duplex connection on the specified socket to be disabled. When `shutdown` is executed, the specified socket is marked unable to send or receive, according to the value of `how`.

- If `how` = 0, the specified socket can no longer receive data. The connection is not completely down until both sides have done a `shutdown` or a `close`.
- If `how` = 1, `shutdown` starts a graceful disconnect by attempting to send any unsent data before preventing further sending. `shutdown` sends an end-of-file condition to the peer, indicating that there are no more data to be sent.

  Once both `shutdown(s, 0)` and `shutdown(s, 1)` have been executed on the same socket descriptor, the only valid operation on the socket at this point is a `close`.
- If `how` = 2, the specified socket can no longer send or receive data. The only valid operation on the socket is a `close`. This has the same effect as executing `shutdown(s, 0)` and `shutdown(s, 1)` on the same socket descriptor.

If you use `close` on a socket, `close` pays attention to the SO_LINGER option, but `shutdown(s, 2)` does not. With `close`, the socket descriptor is deallocated and the last process using the socket destroys it. `shutdown` and its parameters are described in the following table.

Include files:	none
System call:	`shutdown(s,how)` `int s, how;`

Parameter	Description of Contents	INPUT Value
s	socket descriptor	socket descriptor of socket to be shut down
how	number that indicates the type of shutdown	0, 1 or 2

Function result: 0 if shutdown is successful, –1 if failure occurs.

Example: `shutdown (s, 1);`

When to Shut Down a Socket

Optionally, after the process has sent all messages and wants to indicate that it is done sending, shut down the server or client process. Refer to the `shutdown(2)` man page for more information on `shutdown`.

Using Read and Write to Make Stream Sockets Transparent

An example application of `read` and `write` with stream sockets is to fork a command with a socket descriptor as `stdout`. The peer process can `read` input from the command. The command can be any command and does not have to know that `stdout` is a socket. It might use `printf`, which results in the use of `write`. Thus, the stream sockets are transparent.

Sending and Receiving Out-of-band Data

This option is not supported for UNIX Domain (AF_UNIX) sockets.

If an abnormal condition occurs when a process is in the middle of sending a long stream of data, it is useful to be able to alert the other process with an urgent message. The TCP stream socket implementation includes an out-of-band data facility. Out-of-band data uses a **logically** independent transmission channel associated with a pair of connected stream sockets. TCP supports the reliable delivery of only one out-of-band message at a time. The message can be a maximum of one byte long.

Out-of-band data arrive at the destination node in sequence and in stream, but are delivered independently of normal data. If the receiver has enabled the signalling of out-of-band data via the SIOCSPGRP socket `ioctl` (see the `socket(7)` man page), then a SIGURG is delivered when out-of-band data arrive. If the receiver is selecting for exceptional conditions on the receiving socket, it will return true to signal the arrival of out-of-band data. The receiving process can read the out-of-band message and take the appropriate action based on the message contents. A logical mark is placed in the normal data stream to indicate the point at which the out-of-band data were sent, so that data before the message can be handled differently from data following the message. Here is a data stream with an out-of-band marker:

byte stream <----- data	oob mark	data <--------

For a program to know when out-of-band data are available to be received, you may arrange the program to catch the SIGURG signal as follows:

```
struct sigvec vec;
     int onurg();
     int pid, s;

     /*
     ** arrange for onurg() to be called when SIGURG is received:
     */
     vec.sv_handler = onurg;
     vec.sv_mask = 0;
```

```
    vec.sv_onstack = 0;
    if (sigvector(SIGURG, &vec, (struct sigvec *) 0) < 0) {
        perror("sigvector(SIGURG)");
    }
```

onurg() is a routine that handles out-of-band data in the client program.

In addition, the socket's process group must be set, as shown below. The kernel will not send the signal to the process (or process group) unless this is done, even though the signal handler has been enabled. Refer to the socket(7) man page for more details.

```
/*
    ** arrange for the current process to receive SIGURG
    ** when the socket s has urgent data:
    */
    pid = getpid();
    if (ioctl(s, SIOCSPGRP, (char *) &pid) < 0) {
        perror("ioctl(SIOCSPGRP)");
    }
/*
    ** If a process needs to be notified, it should be
    ** pid = -getpgrp();
    */
```

If the server process is sending data to the client process, and a problem occurs, the server can send an out-of-band data byte by executing a send with the MSG_OOB flag set. This sends the out-of-band data and a SIGURG signal to the receiving process.

```
send(sd, &msg, 1, MSG_OOB)
```

When a SIGURG signal is received, onurg is called. onurg receives the out-of-band data byte with the MSG_OOB flag set on a recv call.

It is possible that the out-of-band byte has not arrived when the SIGURG signal arrives. recv *never blocks* on a receive of out-of-band data, so the client may need to repeat the recv call until the out-of-band byte arrives. recv will return EINVAL if the out-of-band data is not available.

Generally, the out-of-band data byte is stored independently from the normal data stream. If, however, the OOB_INLINE socket option has been turned on for this socket, the out-of-band data will remain inline and must be used without the MSG_OOB flag set on a recv() call.

You cannot read *past* the out-of-band pointer location in one recv call. If you request more data than the amount queued on the socket before the out-of-band pointer, then recv will return only the data up to the

out-of-band pointer. However, once you read past the out-of-band pointer location with subsequent `recv` calls, the out-of-band byte can no longer be read.

Usually the out-of-band data message indicates that all data currently in the stream can be flushed. This involves moving the stream pointer with successive `recv` calls, to the location of the out-of-band data pointer.

The request SIOCATMARK informs you, as you receive data from the stream, when the stream pointer has reached the out-of-band pointer. If `ioctl` returns a 0, the next `recv` provides data sent by the server prior to transmission of the out-of-band data. `ioctl` returns a 1 when the stream pointer reaches the out-of-band byte pointer. The next `recv` provides data sent by the server after the out-of-band message. The following shows how the SIOCATMARK request can be used in a SIGURG interrupt handler.

```
/*  s is the socket with urgent data  */

     onurg()
     {
         int atmark;
         char mark;
         char flush [100];

         while (1) {
             /*
             ** check whether we have read the stream
             ** up to the OOB mark yet
             */
             if (ioctl(s, SIOCATMARK, &atmark) < 0) {
                 /* if the ioctl failed */
                 perror("ioctl(SIOCATMARK)");
                 return;
             }
             if (atmark) {
                 /* we have read the stream up to the OOB mark */
                 break;
             }
             /*
             ** read the stream data preceding the mark,
             ** only to throw it away
             */
             if (read(s, flush, sizeof(flush)) <= 0) {
                 /* if the read failed */
                 return;
             }
         }
     /*
     ** receive the OOB byte
      */
      recv(s, &mark, 1, MSG_OOB);
```

```
        printf("received %c OOB\n", mark);
        return;
}
```

4 Using Internet Datagram Sockets

This chapter describes communication between processes using internet datagram sockets.

Overview

Internet UDP datagram sockets provide bidirectional flow of data with record boundaries preserved. However, messages are not guaranteed to be reliably delivered. If a message is delivered, there is no guarantee that it is in sequence and unduplicated, but the data in the message are guaranteed to be intact.

Datagram sockets allow you to send and receive messages *without* establishing a connection. Each message includes a destination address. Processes involved in data transfer are not required to have a client-server relationship; the processes can be symmetrical.

Unlike stream sockets, datagram sockets allow you to send to many destinations from one socket, and receive from many sources with one socket. There is no two-process model, although a two-process model is the simplest case of a more general multiprocess model. The terms **server** and **client** are used in this chapter only in the application sense. There is no difference in the calls that must be made by the processes involved in the data transfer.

For example, you might have a name server process that receives host names from clients all over a network. That server process can send host name and internet address combinations back to the clients. This can all be done with one UDP socket.

The simplest two-process case is used in this chapter to describe BSD Sockets using datagram sockets.

The following table lists the steps involved in exchanging data between datagram sockets.

Table 4-1 **Exchanging Data Between Internet Datagram Sockets**

Client Process Activity	System Call Used	Server Process Activity	System Call Used
create a socket	`socket()`	create a socket	`socket()`
bind a socket address	`bind()`	bind a socket address	`bind()`
send message	`sendto()` or `sendmsg()`		
		receive message	`recvfrom()` or `recvmsg()`
		send message	`sendto()` or `sendmsg()`
receive message	`recvfrom()` or `recvmsg()`		

Each of these steps or activities is described in more detail in the following sections. The description of each activity specifies a system call and includes:

- What happens when the system call is used.
- When to make the call.
- What the parameters do.
- How the call interacts with other BSD Sockets system calls.
- Where to find details on the system call.

The datagram socket program examples are at the end of these descriptive sections. You can refer to the example code as you work through the descriptions.

Preparing Address Variables

Before your client process can make a request of the server process, you must establish the correct variables and collect the information that you need about the server process and the service provided.

The server process needs to:

- Declare socket address variables.
- Assign a wildcard address.
- Get the port address of the service that you want to provide.

The client process needs to:

- Declare socket address variables.
- Get the remote server's internet address.
- Get the port address for the service that you want to use.

These activities are described next. Refer to the program example at the end of this chapter to see how these activities work together.

Declaring Socket Address Variables

You need to declare a variable of type struct `sockaddr_in` to use the local socket address for both processes. For example, the following declarations are used in the example client program:

```
struct sockaddr_in myaddr; /* for local socket address  */
struct sockaddr_in servaddr; /* for server socket address  */
```

`sockaddr_in` is a special case of `sockaddr` and is used with the AF_INET addressing domain. Both types are shown in this chapter, but `sockaddr_in` makes it easier to manipulate the internet and port addresses. Some of the BSD Sockets system calls are declared using a pointer to `sockaddr`, but you can also use a pointer to `sockaddr_in`.

The `sockaddr_in` address structure consists of the following fields:

Field	Description
`short sin_family`	Specifies the address family and should always be set to AF_INET.
`u_short sin_port`	Specifies the port address. Assign this field when you bind the port address for the socket or when you get a port address for a specific service.
`struct in_addr sin_addr`	Specifies the internet address. Assign this field when you get the internet address for the remote host.

The server process must bind the port address of the service to its own socket and establish an address structure to store the clients' addresses when they are received with `recvfrom`. The client process does not have to bind a port address for its local socket; the host binds one automatically if one is not already bound. Refer to the `inet(7F)` man page for more information on `sockaddr_in`.

Getting the Remote Host's Network Address

The client process can use `gethostbyname` to obtain the internet address of the host and the length of that address (as the size of struct `inaddr`) from `/etc/hosts`, NIS, or BIND. `gethostbyname` and its parameters are described in the following table.

Include files: `#include <netdb.h>`

System call:

```
struct hostent *gethostbyname(name)
char *name;
```

Parameter	Description of Contents	INPUT Value
name	pointer to a valid node name (null-terminated string)	host name

Function result: pointer to struct hostent containing internet address, NULL pointer (0) if failure occurs.

Example:

```
#include <netdb.h>
struct hostent *hp; /* point to host info for name server host */
...
servaddr.sin_family = AF_INET;
hp = gethostbyname (argv[1]);
servaddr.sin_addr.s_addr = ((struct in_addr *)(hp->h_addr))->s_addr;
```

The argv[1] parameter is the host name specified in the client program command line. Refer to the gethostent(3N) man page for more information on gethostbyname.

Getting the Port Address for the Desired Service

When a client process needs to use a service that is offered by some server process, it must send a message to the server's socket. The client process must know the port address for that socket. If the service is not in /etc/services, you must add it.

getservbyname obtains the port address of the specified service from /etc/services. getservbyname and its parameters are described in the following table.

Include files: `#include <netdb.h>`

System call:

```
struct servent *getservbyname(name, proto)
char *name, *proto;
```

Parameter	Description of Contents	INPUT Value
name	pointer to a valid service name	service name
proto	pointer to the protocol to be used	udp or 0 if UDP is the only protocol for the service

Function result: pointer to struct servent containing port address, NULL pointer (0) if failure occurs.

Example:

```
#include <netdb.h>
struct servent *sp; /* pointer to service info */
...
sp = getservbyname ("example", "udp");
servaddr.sin_port = sp->s_port;
```

When to Get Server's Socket Address

The server process should get the server's socket address before binding. The client process should get the server's socket address before client requests the service from the host. Refer to the `getservent(3N)` man page for more information on `getservbyname`.

Using a Wildcard Local Address

Wildcard addressing simplifies local address binding. When an address is assigned the value of INADDR_ANY, the host interprets the address as any valid address. This means that the server process can receive on a wildcard address and does not have to look up its own internet address.

For example, to bind a specific port address to a socket, but leave the local internet address unspecified, the following source code could be used:

```
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
...
struct sockaddr_in sin;
...
s = socket(AF_INET, SOCK_DGRAM, 0);
sin.sin_family = AF_INET;
sin.sin_addr.s_addr = INADDR_ANY;
sin.sin_port = MYPORT;
bind (s, &sin, sizeof(sin));
```

Writing the Server and Client Processes

This section explains the calls your server and client processes must make.

Creating Sockets

Both processes must call `socket` to create communication endpoints. `socket` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:
```
s = socket(af, type, protocol)
int s, af, type, protocol;
```

Parameter	Description of Contents	INPUT Value
af	address family	AF_INET
type	socket type	SOCK_DGRAM
protocol	underlying protocol to be used	0 (default) or value returned by `getprotobyname`

Function result: socket number (HP-UX file descriptor), –1 if failure occurs.

Example:

```
ls = socket (AF_INET, SOCK_DGRAM, 0);
```

The socket number returned is the socket descriptor for the newly created socket. This number is an HP-UX file descriptor and can be used for reading, writing or any standard file system calls. A socket descriptor is treated like a file descriptor for an open file.

NOTE

To use `write(2)` with a datagram socket, you must declare a default address. Refer to the "Specifying a Default Socket Address" section of the "Advanced Topics for Internet Datagram Sockets" chapter for more information.

When to Create Sockets

The server or client process should create a socket before any other BSD Sockets system calls. Refer to the `socket(2)` man page for more information on `socket`.

Binding Socket Addresses to Datagram Sockets

After each process has created a socket, it must call `bind` to bind a socket address. Until an address is bound, other processes have no way to reference it.

The server process must bind a specific port address to its socket. Otherwise, a client process would not know what port to send requests to for the desired service.

The client process can let the local host bind its local port address. The client does not need to know its own port address, and if the server process needs to send a reply to the client's request, the server can find out the client's port address when it receives with `recvfrom`.

Set up the address structure with a local address before you make a `bind` call. Use the wildcard address so your processes do not have to look up their own internet addresses. `bind` and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <netinet/in.h>
#include <sys/socket.h>
```

System call:

```
bind (s, addr, addrlen)
int s;
struct sockaddr *addr;
int addrlen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket to be bound
addr	socket address	pointer to address to be bound to s
addrlen	length of socket address	size of struct sockaddr_in_address

Function result: 0 if bind is successful, –1 if failure occurs.

Example:

```
struct sockaddr_in myaddr;
...
bind (s, &myaddr, sizeof(struct sockaddr_in));
```

When to Bind Socket Addresses

The client and server process should bind socket addresses after the socket is created and before any other BSD Sockets system calls. Refer to the `bind(2)` man page for more information on `bind`.

Sending and Receiving Messages

The `sendto` and `recvfrom` (or `sendmsg` and `recvmsg`) system calls are usually used to transmit and receive messages.

Sending Messages

Use `sendto` or `sendmsg` to send messages. `sendmsg` allows the send data to be gathered from several buffers.

If you have declared a default address, you can use `send` , `sendto`, or `sendmsg` to send messages. If you use `sendto` or `sendmsg` in this special case, be sure you specify 0 as the address value or an error will occur.

`send` is described in the "Sending Data" section in the "BSD Sockets: Using Internet Stream Sockets" chapter of this guide and in the `send(2)` man page. `sendto` and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <netinet/in.h>
#include <sys/socket.h>
```

System call:

```
count = sendto(s,msg,len,flags,to,tolen) int s
char *msg;
int len, flags;
struct sockaddr *to
int tolen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket sending message
msg	pointer to data buffer	pointer to data to be sent
len	size of data buffer	size of msg

Parameter	Description of Contents	INPUT Value
flags	settings for optional flags	0 (no options are currently supported)
to	address of recipient socket	pointer to the socket address that message should be sent to
tolen	size of to	length of address structure that to points to

Function result: Number of bytes actually sent, –1 in the event of an error.

Example:

```
count = sendto(s,argv[2],strlen(argv[2]),0,servaddr,sizeof(struct
sockaddr_in));
```

The largest message size for this implementation is 32767 bytes.

You should not count on receiving error messages when using datagram sockets. The protocol is unreliable, meaning that messages may or may not reach their destination. However, if a message reaches its destination, the contents of the message are guaranteed to be intact.

If you need reliable message transfer, you must build it into your application programs or resend a message if the expected response does not occur.

When to Send Data

The client or server process should send data after sockets are bound. Refer to the `send(2)` man page for more information on `sendto` and `sendmsg`.

Receiving Messages

Use `recvfrom` or `recvmsg` to receive messages. `recvmsg` allows the read data to be scattered into buffers.

`recv` can also be used if you do not need to know what socket sent the message. However, if you want to send a response to the message, you must know where it came from. Except for the extra information returned by `recvfrom` and `recvmsg`, the three calls are identical.

`recv` is described in the "Receiving Data" section of the "BSD Sockets: Using Internet Stream Sockets" chapter in this guide and in the `recv(2)` man page. `recvfrom` and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <netinet/in.h>
#include <sys/socket.h>
```

System call:

```
count = recvfrom(s,buf,len,flags,from,fromlen) int
s;
char *buf;
int len, flags;
struct sockaddr *from
int *fromlen;
```

Parameter	Contents	INPUT Value	OUTPUT Value
s	socket descriptor of local socket	socket descriptor of socket receiving message	unchanged
buf	pointer to data buffer	pointer to buffer that is to receive data	pointer to received data
len	maximum number of bytes that should be received	size of data buffer	unchanged

Parameter	Contents	INPUT Value	OUTPUT Value
flags	settings for optional flags	0 or MSG_PEEK	unchanged
from	address of socket that sent message	pointer to address structure, not used for input	pointer to socket address of socket that sent the message
fromlen	pointer to the size of from	pointer to size of from	pointer to the actual size of address returned

Function result: Number of bytes actually received, –1 if an error occurs.

Example:

```
addrlen = sizeof(sockaddr_in);
...
count = recvfrom(s, buffer, BUFFERSIZE, 0, clientaddr, &addrlen);
```

`recvfrom` blocks until there is a message to be received.

No more than `len` bytes of data are returned. The entire message is read in one `recvfrom`, `recvmsg`, `recv` or `read` operation. If the message is too long for the allocated buffer, the excess data are discarded. Because only one message can be returned in a `recvfrom` call, if a second message is in the queue, it is not affected. Therefore, the best technique is to receive as much as possible on each call.

The host does not wait for `len` bytes to be available; if less than `len` bytes are available, that number of bytes are returned.

Flag Options

The flag options are:

- 0 for no options.

- MSG_PEEK for a nondestructive read.

Use the MSG_PEEK option to preview an incoming message. If this option is set on a `recvfrom`, any message returned remains in the data buffer as though it had not been read yet. The next `recvfrom` will return the same message.

When to Receive Data

The client or server process should receive data after sockets are bound. Refer to the `recv(2)` man page for more information on `recvfrom` and `recvmsg`.

Closing a Socket

In most applications, you do not have to worry about cleaning up your sockets. When you exit your program and your process terminates, the sockets are closed for you.

If you need to close a socket while your program is still running, use the HP-UX file system call `close`.

You may have more than one process with the same socket descriptor if the process with the socket descriptor executes a `fork`. `close` decrements the file descriptor count and the calling process can no longer use that file descriptor. When the last `close` is executed on a socket, any unsent messages are sent and the socket is closed. Then the socket is destroyed and can no longer be used.

For syntax and details on `close`, refer to the `close(2)` man page.

Example Using Datagram Sockets

NOTE These programs are provided as examples only of datagram socket usage and are not Hewlett-Packard supported products.

These program examples demonstrate how to set up and use datagram sockets. These sample programs can be found in the `/usr/lib/demos/networking/socket` directory. The client program is intended to run in conjunction with the server program.

This example implements a simple name server. The server process receives requests from the client process. It determines the internet address of the specified host and sends that address to the client process. If the specified host's internet address is unknown, the server process returns an address of all 1s.

The client process requests the internet address of a host and receives the results from the server process.

Before you run the example programs, make the following entry in the two hosts `/etc/services` files:

```
example   22375/udp
```

The source code for these two programs follows.

```
/*                        S E R V . U D P
 *
 * This is an example program that demonstrates the use of
 * datagram sockets as an BSD Sockets mechanism.  This contains
 * the server, and is intended to operate in conjunction with the
 * client program found in client.udp. Together, these two
 * programs demonstrate many of the features of sockets, as well
 * a good conventions for using these features. NOTE: This example
 * is valid only if the /etc/hosts file is being used to lookup
 * host names.
 *
 * This program provides a service called "example".  It is an
 * example of a simple name server.  In order for
 * it to function, an entry for it needs to exist in the
 * /etc/services file.  The port address for this service can be
 * any port number that is likely to be unused, such as 22375,
 * for example.  The host on which the client will be running
 * must also have the same entry (same port number) in its
 * /etc/services file.
 *
 */

#include <sys/types.h>
```

```
#include <sys/socket.h>
#include <netinet/in.h>
#include <stdio.h>
#include <netdb.h>

int s;                    /* socket descriptor */

#define BUFFERSIZE  1024  /* max size of packets to be received */
int cc;                   /* contains the number of bytes read */
char buffer[BUFFERSIZE];  /* buffer for packets to be read into */

struct hostent *hp;       /* pointer to info of requested host */
struct servent *sp;       /* pointer to service information */

struct sockaddr_in myaddr_in;     /* for local socket address */
struct sockaddr_in clientaddr_in;/* for client's socket address */
struct in_addr reqaddr;          /* for requested host's address */

#define ADDRNOTFOUND 0xffffffff /* return address for unfound host */

/*
 *                       M A I N
 *
 * This routine starts the server.  It forks, leaving the child
 * to do all the work, so it does not have to be run in the
 * background.  It sets up the socket, and for each incoming
 * request, it returns an answer.  Each request consists of a
 * host name for which the requester desires to know the
 * internet address.  The server will look up the name in its
 * /etc/hosts file, and return the internet address to the
 * client. An internet address value of all ones will be returned
 * if the host name is not found. NOTE: This example is valid
 * only if the /etc/hosts file is being used to lookup host names.
 *
 */
main(argc, argv)
int argc;
char *argv[];
{
        int addrlen;

                /* clear out address structures */
   memset ((char *)&myaddr_in, 0, sizeof(struct sockaddr_in));
   memset ((char *)&clientaddr_in, 0, sizeof(struct sockaddr_in));

                /* Set up address structure for the socket. */
   myaddr_in.sin_family = AF_INET;
                /* The server should receive on the wildcard address,
                 * rather than its own internet address.  This is
                 * generally good practice for servers, because on
                 * systems which are connected to more than one
                 * network at once will be able to have one server
                 * listening on all networks at once.  Even when the
                 * host is connected to only one network, this is good
                 * practice, because it makes the server program more
                 * portable.
```

```
                  */
myaddr_in.sin_addr.s_addr = INADDR_ANY;
                /* Find the information for the "example" server
                 * in order to get the needed port number.
                 */
sp = getservbyname ("example", "udp");
if (sp == NULL) {
                printf("%s: host not found",
                                argv[0]);
                exit(1);
}
myaddr_in.sin_port = sp->s_port;

                /* Create the socket. */
s = socket (AF_INET, SOCK_DGRAM, 0);
if (s == -1) {
                perror(argv[0]);
                printf("%s: unable to create socket\n", argv[0]);
                exit(1);
}               /* Bind the server's address to the socket. */
if (bind(s, &myaddr_in, sizeof(struct sockaddr_in)) == -1) {
                perror(argv[0]);
                printf("%s: unable to bind address\n", argv[0]);
                exit(1);
}

         /* Now, all the initialization of the server is
          * complete, and any user errors will have already
          * been detected.  Now we can fork the daemon and
          * return to the user.  We need to do a setpgrp
          * so that the daemon will no longer be associated
          * with the user's control terminal.  This is done
          * before the fork, so that the child will not be
          * a process group leader.  Otherwise, if the child
          * were to open a terminal, it would become associated
          * with that terminal as its control terminal.  It is
          * always best for the parent to do the setpgrp.
          */
setpgrp();

switch (fork()) {
case -1: /* Unable to fork, for some reason. */
         perror(argv[0]);
         printf("%s: unable to fork daemon\n", argv[0]);
         exit(1);

case 0:  /* The child process (daemon) comes here. */
         /* Close stdin, stdout, and stderr so they will
          * not be kept open.  From now on, the daemon will
          * not report any error messages.  This daemon
          * will loop forever, waiting for requests and
          * responding to them.
          */
      fclose(stdin);
      fclose(stdout);
      fclose(stderr);
```

```
        /* This will open the /etc/hosts file and keep
         * it open.  This will make accesses to it faster.
         * If the host has been configured to use the NIS
         * server or name server (BIND), it is desirable
         * not to call sethostent(1), because a STREAM
         * socket is used instead of datagrams for each
         * call to gethostbyname().
         */
sethostent(1);
for(;;) {
                /* Note that addrlen passed as a pointer
                 * so that the recvfrom call can return
                 * the size of the returned address.
                 */
        addrlen = sizeof(struct sockaddr_in);

                /* This call will block until a new
                 * request arrives.  Then, it will
                 * return the address of the client,
                 * and a buffer containing its request.
                 * BUFFERSIZE - 1 bytes are read so that
                 * room is left at the end of the buffer
                 * for a null character.
                 */
        cc = recvfrom(s, buffer, BUFFERSIZE - 1, 0,
                        &clientaddr_in, &addrlen);
        if ( cc == -1) exit(1);
                        /* Make sure the message received is
                         * null terminated.
                         */
        buffer[cc]='\0';
                        /* Treat the message as a string
                         * containing a hostname.  Search
                         * fot the name in /etc/hosts.
                         */
        hp = gethostbyname (buffer);
        if (hp == NULL) {
                                /* Name was not found. Return
                                 * a special value signifying
                                 * the error.
                                 */
                reqaddr.s_addr = ADDRNOTFOUND;
        } else {
                                    /* Copy address of host
                                    * into the return buffer.
                                    */
                reqaddr.s_addr =
              ((struct in_addr *)(hp->h_addr))->s_addr;
        }
                        /* Send the response back to the
                         * requesting client. The address is
                         * sent in network byte order. All
                         * errors are ignored. The client
                         * will retry if it does not receive
                         * the response.
                         */
```

```
                        sendto (s, &reqaddr, sizeof(struct in_addr),
                                        0, &clientaddr_in, addrlen);
                }

        default:                /* Parent process comes here. */
                exit(0);
        }
}

/*
 *                      C L I E N T . U D P
 *
 * This is an example program that demonstrates the use of
 * datagram sockets as an BSD Sockets mechanism. This contains
 * the client, and is intended to operate in conjunction with the
 * server program found in serv.udp. Together, these two programs
 * demonstrate many of the features of sockets, as well as good
 * conventions for using these features.
 *
 * This program requests a service called "example".  In order for
 * it to function, an entry for it needs to exist in the
 * /etc/services file.  The port address for this service can be
 * any port number that is likely to be unused, such as 22375,
 * for example.  The host on which the server will be runnin
 * must also have the same entry (same port number) in its
 * /etc/services file.
 *
 * The "example" service is an example of a simple name server
 * application.  The host that is to provide this service is
 * required to be in the /etc/hosts file. Also, the host providing
 * this service presumably knows the internet addresses of many
 * hosts which the local host does not.  Therefore, this program
 * will request the internet address of a target host by name from
 * the serving host.  The serving host will return the requested
 * internet address as a response, and will return an address of
 * all ones if it does not recognize the host name.
 *
 */

#include <sys/types.h>
#include <sys/socket.h>
#include <sys/errno.h>
#include <netinet/in.h>
#include <stdio.h>
#include <signal.h>
#include <netdb.h>

extern int errno;

int s;                          /* socket descriptor */

struct hostent *hp;             /* pointer to info for nameserver host */
struct servent *sp;             /* pointer to service information */

struct sockaddr_in myaddr_in;   /* for local socket address */
struct sockaddr_in servaddr_in;/* for server socket addres */
```

```
struct in_addr reqaddr;         /* for returned internet address */
#define ADDRNOTFOUND 0xffffffff /* value returned for unknown host */
#define RETRIES 5 /* # of times to retry before giving up */

/*
 *                      H A N D L E R
 *
 *      This routine is the signal handler for the alarm signal.
 *      It simply re-installs itself as the handler and returns.
 */
handler()
{
        signal(SIGALRM, handler);
}

/*
 *                      M A I N
 *
 * This routine is the client which requests service from
 * the remote "example server".  It will send a message to the
 * remote nameserver requesting the internet address corresponding
 * to a given hostname. The server will look up the name, and
 * return its internet address. The returned address will be
 * written to stdout.
 *
 * The name of the system to which the requests will be sent is
 * given as the first parameter to the command.  The second
 * parameter should be the name of the target host for which the
 * internet address is sought.
 */
main(argc, argv)
int argc;
char *argv[];
{
        int i;
        int retry = RETRIES;            /* holds the retry count */
        char *inet_ntoa();

        if (argc != 3) {
            fprintf(stderr, "Usage:  %s <nameserver> <target>\n",
                    argv[0]);
            exit(1);
        }

            /* clear out address structures */
        memset ((char *)&myaddr_in, 0, sizeof(struct sockaddr_in));
        memset ((char *)&servaddr_in, 0, sizeof(struct sockaddr_in));
            /* Set up the server address. */
        servaddr_in.sin_family = AF_INET;
            /* Get the host info for the server's hostname that the
             * user passed in.
             */
        hp = gethostbyname (argv[1]);
        if (hp == NULL) {
                fprintf(stderr, "%s: %s not found in /etc/hosts\n",
                                argv[0], argv[1]);
```

```
                exit(1);
        }
        servaddr_in.sin_addr.s_addr = ((struct in_addr *)
                                      (hp->h_addr))->s_addr;
                /* Find the information for the "example" server
                 * in order to get the needed port number.
                 */
        sp = getservbyname ("example", "udp");
        if (sp == NULL) {
            fprintf(stderr, "%s: example not found in /etc/services\n",
                                argv[0]);
                exit(1);
        }
        servaddr_in.sin_port = sp->s_port;

                /* Create the socket. */
        s = socket (AF_INET, SOCK_DGRAM, 0);
        if (s == -1) {
                perror(argv[0]);
                fprintf(stderr, "%s: unable to create socket\n", argv[0]);
                exit(1);
        }
               /* Bind socket to some local address so that the
                * server can send the reply back.  A port number
                * of zero will be used so that the system will
                * assign any available port number.  An address
                * of INADDR_ANY will be used so we do not have to
                * look up the internet address of the local host.
                */
        myaddr_in.sin_family = AF_INET;
        myaddr_in.sin_port = 0;
        myaddr_in.sin_addr.s_addr = INADDR_ANY;
        if (bind(s, &myaddr_in, sizeof(struct sockaddr_in)) == -1) {
                perror(argv[0]);
                fprintf(stderr, "%s: unable to bind socket\n", argv[0]);
                exit(1);
        }
                /* Set up alarm signal handler. */
        signal(SIGALRM, handler);
                /* Send the request to the nameserver. */
again:  if (sendto (s, argv[2], strlen(argv[2]), 0, &servaddr_in,
                                sizeof(struct sockaddr_in)) == -1) {
                perror(argv[0]);
                fprintf(stderr, "%s: unable to send request\n", argv[0]);
                exit(1);
        }
                /* Set up a timeout so I don't hang in case the packet
                 * gets lost.  After all, UDP does not guarantee
                 * delivery.
                 */
        alarm(5);
                /* Wait for the reply to come in.  We assume that
                 * no messages will come from any other source,
                 * so that we do not need to do a recvfrom nor
                 * check the responder's address.
                 */
```

```
        if (recv (s, &reqaddr, sizeof(struct in_addr), 0) == -1) {
                if (errno == EINTR) {
                        /* Alarm went off & aborted the receive.
                         * Need to retry the request if we have
                         * not already exceeded the retry limit.
                         */
                        if (—retry) {
                                goto again;
                        } else {
                                printf("Unable to get response from");
                                printf(" %s after %d attempts.\n",
                                                argv[1], RETRIES);
                                exit(1);
                        }
                } else {
                        perror(argv[0]);
                        fprintf(stderr, "%s: unable to receive response\n",
                                                        argv[0]);
                        exit(1);
                        }
                }
                alarm(0);
                        /* Print out response. */
                if (reqaddr.s_addr == ADDRNOTFOUND) {
                        printf("Host %s unknown by nameserver %s.\n", argv[2],
                                                argv[1]);
                        exit(1);
                } else {
                        printf("Address for %s is %s.\n", argv[2],
                          inet_ntoa(reqaddr));
                }
}
```

5 Advanced Topics for Internet Datagram Sockets

This chapter explains the following:

- SO_BROADCAST socket option.
- Specifying a default socket address.
- Synchronous I/O multiplexing with select.
- Sending and receiving data asynchronously.
- Sending and receiving IP multicast datagrams.
- Nonblocking I/O.
- Using broadcast addresses.

SO_BROADCAST Socket Option

This option is AF_INET socket-specific.

SO_BROADCASTADDR establishes permission to send broadcast datagrams from the socket.

Specifying a Default Socket Address

It is possible (but not required) to specify a default address for a remote datagram socket.

This allows you to send messages without specifying the remote address each time. In fact, if you use `sendto` or `sendmsg`, an error occurs if you enter any value other than 0 for the socket address after the default address has been recorded. You can use `send` or `write` instead of `sendto` or `sendmsg` once you have specified the default address.

Use `recv` for receiving messages. Although `recvfrom` can be used, it is not necessary, because you already know that the message came from the default remote socket. (Messages from sockets other than the default socket are discarded without notice.) `read(2)` can also be used, but does not allow you to use the MSG_PEEK flag.

Specify the default address with the `connect` system call. `connect` recognizes two special default addresses, INADDR_ANY and INADDR_BROADCAST. Using INADDR_ANY connects your socket to the IP address of your local host's primary LAN interface (for loopback connections). Using INADDR_BROADCAST connects your socket to the subnet broadcast address for your primary LAN interface; it allows you to send out broadcast packets that interface without specifying the subnet broadcast address.

When a datagram socket descriptor is specified in a `connect` call, `connect` associates the specified socket with a particular remote socket address. `connect` returns immediately because it only records the peer's socket address. After `connect` records the default address, any message sent from that socket is automatically addressed to the peer process and only messages from that peer are delivered to the socket.

`connect` may be called any number of times to change the associated destination address.

NOTE This call does not behave the same as a `connect` for stream sockets. There is no connection, just a default destination. The remote host that you specify as the default may or may not use `connect` to specify your local host as its default remote host. The default remote host is **not** notified if your local socket is destroyed.

`connect` and its parameters are described in the following table.

Include files:
```
#include<sys/types.h>
#include<netinet/in.h>
#include<sys/socket.h>
```

System call:
```
connect(s,addr,addrlen)
int s;
struct sockaddr*addr;
int addrlen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket requesting default peer address
addr	pointer to the socket address	pointer to socket address of the socket to be the peer
addrlen	length of address	length of address pointed to by addr

Function result: 0 if connect is successful, –1 if failure occurs.

When to Specify a Default Socket Address

The client or server process should specify a default socket address after sockets are bound.

Synchronous I/O Multiplexing with Select

The `select` system call can be used with sockets to provide a synchronous multiplexing mechanism. The system call has several parameters which govern its behavior. If you specify a zero pointer for the timeout parameter **timout**, `select` will block until one or more of the specified socket descriptors is ready. If **timout** is a non-zero pointer, it specifies a maximum interval to wait for the selection to complete.

`select` is useful for datagram socket descriptors to determine when data have arrived and are ready to be read without blocking; use the FIONREAD parameter to the `ioctl` system call to determine exactly how much data are available.

`select` for exceptional conditions will return true for BSD sockets if out-of- band data is available. `select` will always return true for sockets which are no longer capable of being used (e.g. if a `close` or `shutdown` system call has been executed against them).

`select` is used in the same way as in other applications. Refer to the `select(2)` man page for information on how to use `select`.

Sending and Receiving Data Asynchronously

Asynchronous sockets allow a user program to receive an SIGIO signal when the state of the socket changes. This state change can occur, for example, when new data arrive. More information on SIGIO can be found in the "Advanced Topics for Internet Stream Sockets" chapter of this guide.

Sending and Receiving IP Multicast Datagrams

IP multicasting provides a mechanism for sending a single datagram to a group of systems. Normally, only systems that have joined the multicast group process the datagrams.

Multicast datagrams are transmitted and delivered with the same "best effort" reliability as regular unicast IP datagrams. That is, the datagrams are not guaranteed to arrive intact at all members of the destination group or in the same order as the datagrams were sent.

Membership in a multicast group is dynamic. That is, systems can join and leave groups at any time. A system remains a member of a multicast group until the last socket that joined the group is closed or has dropped membership in the group. A system can be a member of more than one group at a time. A system that has multiple interfaces might be a member of the same group on each interface.

Sending IP Multicast Datagrams

IP multicasting is supported only for AF_INET sockets of type SOCK_DGRAM and only on networks for which the interface supports multicasting.

To send a multicast datagram, the application specifies an IP multicast address as the destination address in a `sendto(2)` call. For example:

```
#include <netinet/in.h>
struct sockaddr_in servaddr;
servaddr.sin_family = AF_INET;
servaddr.sin_port = 12345;
servaddr.sin_addr.s_addr = inet_addr("224.1.2.3");
sendto(s, buf, buflen, 0, &servaddr, sizeof(servaddr));
```

It is not necessary for the system to join a multicast group in order to send multicast datagrams.

Each multicast datagram is sent from one network interface at a time, even if the system has joined the destination multicast group on more than one interface. The system administrator configures a default interface from which to send multicast datagrams. An application can override the default by setting a socket option IP_MULTICAST_IF to specify the outbound interface for datagrams sent through that socket.

Normally, multicast datagrams are sent only to systems directly connected to the same network that the sending interface is on. If multicast datagrams are intended for distant networks, a special multicast router must be present on the local and intermediate networks. A socket option IP_MULTICAST_TTL controls the number of intermediate systems through which a multicast datagram can be forwarded.

If a multicast datagram is sent to a port from which an application on the local system is reading, normally a copy of the datagram is looped back and delivered to the application. A socket option IP_MULTICAST_LOOP allows the sending application to disable loopback for datagrams sent through that socket.

Specifying the Outbound Interface IP_MULTICAST_IF

Normally, multicast datagrams are sent through the interface that is associated with the default route, if one is configured. Alternatively, the system administrator can configure other interfaces as the default multicast route and as the route for specific multicast groups.

The configured multicast routes should suffice for most applications. However, an application can override the default multicast route by setting the IP_MULTICAST_IF socket option. For example:

```
#include <netinet/in.h>
struct in_addr addr;
addr.s_addr = inet_addr("192.1.2.3");
setsockopt(s, IPPROTO_IP, IP_MULTICAST_IF, &addr, sizeof(addr));
```

If `addr.s_addr` is INADDR_ANY, subsequent outbound multicasts from the application are sent from the default interface. Otherwise, `addr.s_addr` must specify the IP address of a local network interface that supports multicasting.

Specifying the Scope of a Multicast IP_MULTICAST_TTL

By default, multicast datagrams are sent only to systems on a local network. If the multicast datagram is intended to be sent to distant networks, and if there is a special multicast router on the local network, the application can increase the scope of the multicast by using the IP_MULTICAST_TTL socket option. For example:

```
#include <netinet/in.h>
unsigned char ttl = 64;
setsockopt(s, IPPROTO_IP, IP_MULTICAST_TTL, &ttl, sizeof(ttl));
```

Note that `ttl` is an unsigned char. Any of the values 0 through 255 can be specified. If `ttl` is zero, the multicast is limited to the local system. If `ttl` is one, the multicast is limited to a local network. If `ttl` is two, the multicast can forwarded through at most one gateway; and so forth.

Disabling Loopback IP_MULTICAST_LOOP

Normally, if a multicast datagram is sent to a port from which an application on the local system is reading, a copy of the datagram is looped back and delivered to the application.

The IP_MULTICAST_LOOP socket option allows the sending application to disable loopback for datagrams sent through that socket. For example:

```
#include <netinet/in.h>
unsigned char loop = 0;
setsockopt(s, IPPROTO_IP, IP_MULTICAST_LOOP, &loop, sizeof(loop));
```

Note that `loop` is an unsigned char. Usually, applications should not disable loopback. Disabling loopback provides only a modest improvement in system performance.

Receiving IP Multicast Datagrams

IP multicasting is supported only for AF_INET sockets of type SOCK_DGRAM and only on networks for which the interface supports multicasting.

In order to receive multicast datagrams, an application must bind to the port number to which multicast datagrams will be sent. If an application binds to the address INADDR_ANY, it may receive all datagrams that are sent to the port number. If the application binds to a multicast group address, it may receive only datagrams sent to that group and port number.

Additionally, the system must join the multicast group on the interface on which the multicast datagrams arrive. An application can request that the system join a group by using the IP_ADD_MEMBERSHIP socket option.

An application can join up to IP_MAX_MEMBERSHIPS multicast groups on each socket. Currently, this is defined to be 20 in `<netinet/in.h>`. However, each network interface may impose a smaller system-wide limit because of interface resource limitations and

because the system uses some link-layer multicast addresses. For example, the E/ISA interface card is limited to 16 multicast addresses, and the system uses two of those. So all applications in the system can join at most 14 unique multicast groups on each E/ISA interface.

An application automatically leaves a multicast group when it terminates, either normally or abnormally. Alternatively, the application can leave a group by using the IP_DROP_MEMBERSHIP socket option. However, the system remains a member of a multicast group as long as at least one application is a member.

If more than one application binds to the same port number on a system, each application must set the SO_REUSEPORT socket option *before* binding to the port. In that case, every application will receive all multicast datagrams sent to the port number.

Joining a Multicast Group IP_ADD_MEMBERSHIP

In order to receive multicast datagrams, a system must join the multicast group. An application can request that the system join a group by using the IP_ADD_MEMBERSHIP socket option. For example:

```
#include <netinet/in.h>
struct ip_mreq mreq;
mreq.imr_multiaddr.s_addr = inet_addr("224.1.2.3");
mreq.imr_interface.s_addr = INADDR_ANY;
setsockopt(s, IPPROTO_IP, IP_ADD_MEMBERSHIP, &mreq, sizeof(mreq));
```

Each membership is associated with only one interface, and it is possible to join the same multicast group on more than one interface. If `imr_interface` is INADDR_ANY, the membership is joined on the system default interface. Otherwise, `imr_interface` should be the IP address of a local interface.

Leaving a Multicast Group IP_DROP_MEMBERSHIP

An application automatically leaves a multicast group when it terminates, either normally or abnormally. Alternatively, the application can leave a group by using the IP_DROP_MEMBERSHIP socket option. For example:

```
#include <netinet/in.h>
struct ip_mreq mreq;
mreq.imr_multiaddr.s_addr = net_addr("224.1.2.3");
mreq.imr_interface.s_addr = INADDR_ANY;
setsockopt (s, IPPROTO_IP, IP_DROP_MEMBERSHIP, &mreq, sizeof(mreq));
```

Note that `imr_interface` must match the field that was used when the IP_ADD_MEMBERSHIP socket option was specified for `imr_multiaddr`.

Sharing a Multicast Port SO_REUSEPORT

If more than one application may bind to the same port number on a system, each application must set the SO_REUSEPORT socket option *before* binding to the port. For example:

```
#include <netinet/in.h>
int reuse = 1;
setsockopt(s, SOL_SOCKET, SO_REUSEPORT, &reuse, sizeof(reuse));
```

Nonblocking I/O

Sockets are created in blocking mode I/O by default. You can specify that a socket be put in nonblocking mode by using the `ioctl` system call with the FIOSNBIO request.

An example usage of this call is:

```
#include<sys/ioctl.h>
...
ioctl(s,FIOSNBIO,&arg);
```

`arg` is a pointer to `int`:

- When `int` equals 0, the socket is changed to blocking mode.
- When `int` equals 1, the socket is changed to nonblocking mode.

If a socket is in nonblocking mode, the following calls are affected:

`recvfrom` — If no messages are available to be received, `recvfrom` returns the value -1 and the EWOULDBLOCK error. This is also true for `recv` and `read`.

`sendto` — If there is no available message space for the message to be transmitted, `sendto` returns the value -1 and the EWOULDBLOCK error.

The O_NDELAY flag for `fcntl(2)` is also supported. If you use this flag and there is no message available to be received on a `recv`, `recvfrom`, or `read` call, the call returns immediately with the value of 0. If you use the O_NONBLOCK flag, the call returns immediately with the value of -1 and the EAGAIN error. This is the same as returning an end-of-file condition. This is also true for `send` , `sendto`, and `write` if there is not enough buffer space to complete the send.

The O_NDELAY and O_NONBLOCK flags have precedence over the FIOSNBIO flag. Setting both the O_DELAY and O_NONBLOCK flags is not allowed.

Using Broadcast Addresses

In place of a unique internet address or the wildcard address, you can also specify a broadcast address. A broadcast address is an internet address with a local address portion of all 1s.

If you use broadcast addressing, be careful not to overload your network.

6 Using UNIX Domain Stream Sockets

This chapter describes creating a UNIX Domain stream socket connection between two processes executing on the same node.

Overview

UNIX Domain (AF_UNIX) stream sockets provide bidirectional, reliable, unduplicated flow of data without record boundaries. They offer significant performance increases when compared with the use of local internet (AF_INET) sockets, due primarily to lower code execution overhead. The following table lists the steps involved in creating and terminating a UNIX Domain BSD Sockets connection using stream sockets.

Table 6-1 **Creating/Terminating BSD Sockets Connections Using UNIX Domain Stream Sockets**

Client Process Activity	System Call Used	Server Process Activity	System Call Used
create a socket	`socket()`	create a socket	`socket()`
		bind a socket address	`bind()`
		listen for incoming connection requests	`listen()`
request a connection	`connect()`		
		accept connection	`accept()`
send data	`write()` or `send()`		
		receive data	`read()` or `recv()`

Client Process Activity	System Call Used	Server Process Activity	System Call Used
		send data	`write()` or `send()`
receive data	`read()` or `recv()`		
disconnect socket (optional)	`shutdown()` or `close()`	disconnect socket (optional)	`shutdown()` or `close()`

Each of these steps or activities is described in more detail in the following sections. The description of each activity specifies a system call and includes:

- What happens when the system call is used.
- When to make the call.
- What the parameters do.
- How the call interacts with other BSD Sockets system calls.
- Where to find details on the system call.

The UNIX Domain stream socket program examples are at the end of these descriptive sections. You can refer to the example code as you work through the descriptions.

Preparing Address Variables

Before you begin to create a connection, establish the correct variables and collect the information that you need to request a connection.

Your server process needs to:

- Declare socket address variables.
- Get the pathname (character string) for the service you want to provide.

Your client process needs to:

- Declare socket address variables.
- Get the pathname (character string) for the service you want to use.

These activities are described next. Refer to the program example at the end of this chapter to see how these activities work together.

Declaring Socket Address Variables

You need to declare a variable of type struct sockaddr_un to use for socket addresses. For example, the following declarations are used in the example client program:

```
struct sockaddr_un myaddr; /* for local socket address   */
struct sockaddr_un peeraddr; /* for peer socket address   */
```

sockaddr_un is a special case of sockaddr and is used with the AF_UNIX address domain .

The sockaddr_un address structure consists of the following fields:

Field	Description
short sun_family	Specifies the address family and should always be set to AF_UNIX.
u_char sun_path[92]	Specifies the pathname to which the socket is bound or will be bound (e.g. /tmp/mysocket).

The server process only needs an address for its own socket. Your client process will not need an address for its own socket.

Writing the Server Process

This section explains the calls your server process must make to connect with and serve a client process.

Creating a Socket

The server process must call `socket` to create a communication endpoint. `socket` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:
```
s = socket(af, type, protocol) int af,
type, protocol;
```

Parameter	Description of Contents	INPUT Value
af	address family	AF_UNIX
type	socket type	SOCK_STREAM
protocol	underlying protocol to be used	0 (default)

Function result: socket number (HP-UX file descriptor), –1 if failure occurs.

Example: `s = socket (AF_UNIX, SOCK_STREAM, 0);`

The socket number returned is the socket descriptor for the newly created socket. This number is an HP-UX file descriptor and can be used for reading, writing or any standard file system calls after a BSD Sockets connection is established. A socket descriptor is treated like a file descriptor for an open file.

When to Create Sockets

The server process should create sockets before any other BSD Sockets system calls. Refer to the `socket(2)` man page for more information on `socket`.

Binding a Socket Address to the Server Process's Socket

After your server process has created a socket, it must call `bind` to bind a socket address. Until an address is bound to the server socket, other processes have no way to reference it.

The server process must bind a specific pathname to this socket, which is used for listening. Otherwise, a client process would not know what pathname to connect to for the desired service.

Set up the address structure with a local address before you make a `bind` call. `bind` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/un.h>
#include <sys/socket.h>
```

System call:
```
bind (s, addr, addrlen) int s; struct
sockaddr_un *addr; int addrlen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket to be bound
addr	socket address	pointer to address to be bound to s
addrlen	length of socket address	size of struct sockaddr_un

Function result: 0 if bind is successful, −1 if failure occurs.

Example:
```
struct sockaddr_un myaddr;
...
bind (ls, &myaddr, sizeof(struct sockaddr_un));
```

When to Bind Socket Addresses

The server process should bind socket addresses after the socket is created and before any other BSD Sockets system calls. Refer to the `bind(2)` man page for more information on `bind`.

Setting the Server Up to Wait for Connection Requests

Once your server process has an address bound to it, it must call `listen` to set up a queue that accepts incoming connection requests. The server process then monitors the queue for requests (using `select(2)` or `accept`). The server process cannot respond to a connection request until it has executed `listen`. `listen` and its parameters are described in the following table.

Include files: none

System call:

```
listen(s, backlog)
int s, backlog;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	server socket's descriptor
backlog	preferred maximum number of connection requests in the queue at any time	size of queue (between 0 and SOMAXCONN)

Function result: 0 if listen is successful, −1 if failure occurs.

Example: `listen (ls, 5);`

`backlog` is the preferred number of unaccepted incoming connections allowed at a given time. The actual number may be greater than the specified `backlog`. When the queue is full, further connection requests are rejected.

A `backlog` of 0 specifies only 1 pending connection can exist at any given time. SOMAXCONN is defined in `<sys/socket.h>`. The current default setting is 20, but may change in future releases.

When to Set Server Up to Listen

The server process should set up server to listen after socket is created and bound and before the server can respond to connection requests. Refer to the `listen(2)` man page for more information on `listen`.

Accepting a Connection

The server process can accept any connection requests that enter its queue after it executes `listen`. `accept` creates a new socket for the connection and returns the socket descriptor for the new socket. The new socket:

- Is created with the same properties as the old socket.
- Has the same bound pathname as the old socket.
- Is connected to the client process' socket.

`accept` blocks until there is a connection request from a client process in the queue, unless you are using nonblocking I/O. `accept` and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <sys/un.h>
#include <sys/socket.h>
```

System call:

```
s = accept(ls,addr,addrlen)
int s;
int ls;
struct sockaddr_un *addr;
int *addrlen;
```

Parameter	Contents	INPUT Value	OUTPUT Value
s	socket descriptor of local socket	socket descriptor of server socket	unchanged
addr	socket address	pointer to address structure where address will be put	pointer to socket address of client socket that server's new socket is connected to
addrlen	length of address	pointer to the size of struct sockaddr_un	pointer to the actual length of address returned in addr

Function result: socket descriptor of new socket if accept is successful, –1 if failure occurs.

Example:

```
struct sockaddr_un peeraddr;
...
addrlen = sizeof(sockaddr_un);
s = accept (ls, &peeraddr, &addrlen);
```

There is no way for the server process to indicate which requests it can accept. It must accept all requests or none.

When to Accept a Connection

The server process should accept a connection after executing the listen call. Refer to the `accept(2)` man page for more information on `accept`.

Writing the Client Process

This section discusses the calls your client process must make to connect with and be served by a server process.

Creating a Socket

The client process must call `socket` to create a communication endpoint. `socket` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:
```
s = socket(af, type, protocol)
int af, type, protocol;
```

Parameter	Description of Contents	INPUT Value
af	address family	AF_UNIX
type	socket type	SOCK_STREAM
protocol	underlying protocol to be used	0 (default)

Function result: socket number (HP-UX file descriptor), –1 if failure occurs.

Example: `s = socket (AF_UNIX, SOCK_STREAM, 0);`

The socket number returned is the socket descriptor for the newly created socket. This number is an HP-UX file descriptor and can be used for reading, writing or any standard file system calls after a BSD Sockets connection is established. A socket descriptor is treated like a file descriptor for an open file.

When to Create Sockets

The client process should create sockets before requesting a connection. Refer to the `socket(2)` man page for more information on `socket`.

Requesting a Connection

Once the server process is listening for connection requests, the client process can request a connection with the `connect` call. `connect` and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <sys/un.h>
#include <sys/socket.h>
```

System call:

```
connect(s, addr, addrlen)
int s;
struct sockaddr_un *addr;
int addrlen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket requesting a connection
addr	pointer to the socket address	pointer to the socket address of the socket to which client wants to connect
addrlen	length of addr	size of address structure pointed to by addr

Function result: 0 if connect is successful, –1 if failure occurs.

Example:

```
struct sockaddr_un peeraddr;
...
connect (s, &peeraddr, sizeof(struct sockaddr_un));
```

`connect` initiates a connection. When the connection is ready, the client process completes its `connect` call and the server process can complete its `accept` call.

NOTE The client process does not get feedback that the server process has completed the `accept` call. As soon as the `connect` call returns, the client process can send data.

When to Request a Connection

The client process should request a connection after socket is created and after server socket has a listening socket. Refer to the `connect(2)` man page for more information on `connect`.

Sending and Receiving Data

After the connect and accept calls are successfully executed, the connection is established and data can be sent and received between the two socket endpoints. Because the stream socket descriptors correspond to HP-UX file descriptors, you can use the read and write calls (in addition to send and recv) to pass data through a socket-terminated channel.

If you are considering the use of the read and write system calls instead of the send and recv calls described below, you should consider the following:

- If you use read and write instead of send and recv, you can use a socket for stdin or stdout.
- If you use read and write instead of send and recv, you cannot use the options specified with the send or recv flags parameter.

See the table that lists other system calls in chapter 8, for more information on which of these system calls are best for your application.

Sending Data

send and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:

```
count = send(s,msg,len,flags)
int s;
char *msg;
int len, flags;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket sending data
msg	pointer to data buffer	pointer to data to be sent
len	size of data buffer	size of msg
flags	settings for optional flags	0

Function result: number of bytes actually sent, –1 if failure occurs.

Example: `count = send (s, buf, 10, 0);`

`send` blocks until the specified number of bytes have been queued to be sent, unless you are using nonblocking I/O.

When to Send Data

The server or client process should send data after connection is established. Refer to the `send(2)` man page for more information on `send`.

Receiving Data

`recv` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:
```
count = recv(s,buf,len,flags)
int s;
char *buf;
int len, flags;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket sending data
buf	pointer to data buffer	pointer to data to be sent
len	maximum number of bytes that should be received	size of data buffer
flags	settings for optional flags	0

Function result: number of bytes actually received, –1 if failure occurs.

Example: `count = recv(s, buf, 10, 0);`

`recv` blocks until there is at least 1 byte of data to be received, unless you are using nonblocking I/O. The host does not wait for `len` bytes to be available; if less than `len` bytes are available, that number of bytes are received.

No more than `len` bytes of data are received. If there are more than `len` bytes of data on the socket, the remaining bytes are received on the next `recv`.

Flag Options

There are no flag options for UNIX Domain (AF_UNIX) sockets. The only supported value for this field is 0.

When to Receive Data

The server or client process should receive data after connection is established. Refer to the `recv(2)` man page for more information on `recv`.

Closing a Socket

In most applications, you do not have to worry about cleaning up your sockets. When you exit your program and your process terminates, the sockets are closed for you.

If you need to close a socket while your program is still running, use the `close` system call. For example, you may have a daemon process that uses `fork` to create the server process. The daemon process creates the BSD Sockets connection and then passes the socket descriptor to the server. You then have more than one process with the same socket descriptor. The daemon process should do a `close` of the socket descriptor to avoid keeping the socket open once the server is through with it. Because the server performs the work, the daemon does not use the socket after the `fork`.

`close` decrements the file descriptor reference count and the calling process can no longer use that file descriptor.

When the last `close` is executed on a socket descriptor, any unsent data are sent before the socket is closed. Any unreceived data are lost.

Example Using UNIX Domain Stream Sockets

NOTE These programs are provided as examples only of UNIX Domain stream socket usage and are not Hewlett-Packard supported products.

These programming examples demonstrate how to set up and use UNIX Domain stream sockets. These sample programs can be found in the `/usr/lib/demos/networking/af_unix` directory. The client program is intended to run in conjunction with the server program.

This example shows how to create UNIX Domain stream sockets and how to set up address structures for the sockets. In this example the client process sends 2000 bytes of data to the server (five times). The server process can receive data from any other process and will echo the data back to the sender.

```
/*
 *      Sample Program: AF_UNIX stream sockets, server process
 *
 *      CATCH - RECEIVE DATA FROM THE PITCHER
 *
 *      Pitch and catch set up a simple unix domain stream socket
 *      client-server connection. The client (pitch) then sends
 *      data to server (catch), throughput is calculated, and the
 *      result is printed to the client's stdout.
 */
#include <stdio.h>
#include <time.h>
#include <signal.h>
#include <errno.h>
#include <sys/types.h>
#include <sys/socket.h>
#include <sys/un.h>
#define SOCKNAME        "/tmp/p_n_c"
#define BUFSIZE         32*1024-1
int     timeout();
int     s;              /* server socket */
char buffer[BUFSIZE];
struct bullet {
     int bytes;
     int throughput;
     int magic;
} bullet = { 0, 0, 0 };

send_data(fd, buf, buflen)
     char *buf;
{
     int cc;
     while (buflen > 0) {
```

```
            cc = send(fd, buf, buflen, 0);
            if (cc == -1) {
                 perror("send");
                 exit(0);
            }
            buf += cc;
            buflen -= cc;
       }
}
recv_data(fd, buf, buflen)
     char *buf;
{
     int cc;
     while (buflen > 0) {
          cc = recv(fd, buf, buflen, 0);
          if (cc == -1) {
               perror("recv");
               exit(0);
          }
          buf += cc;
          buflen -= cc;
     }
}
main(argc, argv)
     int argc;
     char *argv[];
{
     int bufsize, bytes, cc, i, total, pid, counter_pid;
     float msec;
     struct timeval tp1, tp2;
     int ns, recvsize, secs, usec;
     struct timezone tzp;
     struct sockaddr_un sa;
/*
 * The SIGPIPE signal will be received if the peer has gone away
 * and an attempt is made to write data to the peer. Ignoring this
 * signal causes the write operation to receive an EPIPE error.
 * Thus, the user is informed about what happened.
*/
     signal(SIGPIPE, SIG_IGN);
     signal(SIGCLD, SIG_IGN);
     signal(SIGINT, timeout);
     setbuf(stdout, 0);
     setbuf(stderr, 0);
     if (argc > 1) {
          argv++;
          counter_pid = atoi(*argv++);
     } else
          counter_pid = 0;
/*
 * Set up the socket variables - address family, socket name.
 * They'll be used later to bind() the name to the server socket.
 */
     sa.sun_family = AF_UNIX;
     strncpy(sa.sun_path, SOCKNAME,
                    (sizeof(struct sockaddr_un) - sizeof(short)));
/*
 * Create the server socket
 */
     if ((s = socket( AF_UNIX, SOCK_STREAM, 0)) == -1) {
```

```
          perror("catch - socket failed");
          exit(0);
     }
     bufsize = BUFSIZE;
/*
 * Use setsockopt() to change the socket buffer size to improve
 * throughput for large data transfers
 */
     if ((setsockopt(s, SOL_SOCKET, SO_RCVBUF, &bufsize,
sizeof(bufsize)))
          == -1) {
               perror("catch - setsockopt failed");
               exit(0);
     }
/*
 * Bind the server socket to its name
 */
     if ((bind(s, &sa, sizeof(struct sockaddr_un))) == -1) {
          perror("catch - bind failed");
          exit(0);
     }
/*
 * Call listen() to enable reception of connection requests
 * (listen() will silently change given backlog 0, to be 1 instead)
 */
     if ((listen(s, 0)) == -1) {
          perror("catch - listen failed");
          exit(0);
     }next_conn:
     i = sizeof(struct sockaddr_un);
/*
 * Call accept() to accept connection request. This call will
block
 * until a connection request arrives.
 */
     if ((ns = accept(s, &sa, &i)) == -1) {
          if (errno == EINTR)
               goto next_conn;
          perror("catch - accept failed");
          exit(0);
     }
     if ((pid = fork()) != 0) {
          close(ns);
          goto next_conn;
     }
/*
     close(s);
*/
/*
 * Receive the bullet to synchronize with the other side
 */
     recv_data(ns, &bullet, sizeof(struct bullet));
     if (bullet.magic != 12345) {
          printf("catch: bad magic %d\n", bullet.magic);
          exit(0);
     }
     bytes = bullet.bytes;
     recvsize = (bytes>BUFSIZE)?BUFSIZE:bytes;
/*
 * Send the bullet back to complete synchronization
```

```
 */
     send_data(ns, &bullet, sizeof(struct bullet));
     cc = 0;
     if (counter_pid)
          kill(counter_pid, SIGUSR1);
     if (gettimeofday(&tp1, &tzp) == -1) {
          perror("catch time of day failed");
          exit(0);
     }
/*
 * Receive data from the client
 */
     total = 0;
     i = bytes;
     while (i > 0) {
          cc = recvsize < i ? recvsize : i;
          recv_data(ns, buffer, cc);
          total += cc;
          i -= cc;
     }
/*
 * Calculate throughput
 */
     if (gettimeofday(&tp2, &tzp) == -1) {
          perror("catch time of day failed");
          exit(0);
     }
     if (counter_pid)
          kill(counter_pid, SIGUSR2);
     secs = tp2.tv_sec - tp1.tv_sec;
     usec = tp2.tv_usec - tp1.tv_usec;
     if (usec < 0) {
          secs ;
          usec += 1000000;
     }
     msec = 1000*(float)secs;
     msec += (float)usec/1000;
     bullet.throughput = bytes/msec;
/*
 * Send back the bullet with throughput info, then close the
 * server socket
 */
     if ((cc = send(ns, &bullet, sizeof(struct bullet), 0)) == -1)
{
          perror("catch - send end bullet failed");
          exit(0);
     }
     close(ns);
}
timeout()
{
        printf( "alarm went off -- stopping the catch process\n");
        fprintf(stderr, "stopping the catch process\n");
        unlink(SOCKNAME);
        close(s);
        exit(6);
}
/*
*       Sample Program : AF_UNIX stream sockets, client process
*
```

```
*       PITCH - SEND DATA TO THE CATCHER
*
*       Pitch and catch set up a simple unix domain stream socket
*       client-server connection. The client (pitch) then sends
*       data to the server (catch), throughput is calculated, and
*       the result is printed to the client's stdout.
*/
#include <stdio.h>
#include <time.h>
#include <netdb.h>
#include <signal.h>
#include <sys/types.h>
#include <sys/socket.h>
#include <sys/un.h>
#define SOCKNAME    "/tmp/p_n_c"
#define BUFSIZE          32*1024-1
char buffer[BUFSIZE];
struct bullet {
     int bytes;
     int throughput;
     int magic;
} bullet = { 0, 0, 12345 };
send_data(fd, buf, buflen)
     char *buf;
{
     int cc;
     while (buflen > 0) {
          cc = send(fd, buf, buflen, 0);
          if (cc == -1) {
               perror("send");
               exit(0);
          }
          buf += cc;
          buflen -= cc;
     }
}
recv_data(fd, buf, buflen)
     char *buf;
{
     int cc;
     while (buflen > 0) {
          cc = recv(fd, buf, buflen, 0);
          if (cc == -1) {
               perror("recv");
               exit(0);
          }
          buf += cc;
          buflen -= cc;
     }
}
main( argc, argv)
     int argc;
     char *argv[];
{
     int bufsize, bytes, cc, i, total, pid;
     float msec;
     struct timeval tp1, tp2;
     int s, sendsize, secs, usec;
     struct timezone tzp;
     struct sockaddr_un sa;
```

```
/*
 * The SIGPIPE signal will be received if the peer has gone away
 * and an attempt is made to write data to the peer.  Ignoring
 * the signal causes the write operation to receive an EPIPE error.
 * Thus, the user is informed about what happened.
*/
signal(SIGPIPE, SIG_IGN);
     setbuf(stdout, 0);
     setbuf(stderr, 0);
     if (argc < 2) {
          printf("usage: pitch Kbytes [pid]\n");
          exit(0);
     }
     argv++;
/*
 * Set up socket variables (address family; name of server socket)
 * (they'll be used later for the connect() call)
 */
     sa.sun_family = AF_UNIX;
     strncpy(sa.sun_path, SOCKNAME,
                (sizeof(struct sockaddr_un) - sizeof(short)));
     bullet.bytes = bytes = 1024*atoi(*argv++);
     if (argc > 2)
          pid = atoi(*argv++);
     else
          pid = 0;
     sendsize = (bytes < BUFSIZE) ? bytes : BUFSIZE;
/*
 * Create the client socket
 */
     if ((s = socket( AF_UNIX, SOCK_STREAM, 0)) == -1) {
          perror("pitch - socket failed");
          exit(0);
     }
     bufsize = BUFSIZE;
/*
 * Change the default buffer size to improve throughput for
 * large data transfers
 */
     if ((setsockopt(s, SOL_SOCKET, SO_SNDBUF, &bufsize,
sizeof(bufsize)))
          == -1) {
               perror("pitch - setsockopt failed");
               exit(0);
     }
/*
 * Connect to the server
 */
     if ((connect(s, &sa, sizeof(struct sockaddr_un))) == - 1) {
          perror("pitch - connect failed");
          exit(0);
     }
/*
 * send and receive the bullet to synchronize both sides
 */
     send_data(s, &bullet, sizeof(struct bullet));
     recv_data(s, &bullet, sizeof(struct bullet));
     cc = 0;
     if (pid)
          kill(pid,SIGUSR1);
```

```
     if (gettimeofday(&tp1, &tzp) == -1) {
          perror("pitch time of day failed");
          exit(0);
     }
     i = bytes;
     total = 0;
/*
 * Send the data
 */
     while (i > 0) {
          cc = sendsize < i ? sendsize : i;
          send_data(s, buffer, cc);
          i -= cc;
          total += cc;
     }
/*
 * Receive the bullet to calculate throughput
 */
     recv_data(s, &bullet, sizeof(struct bullet));
     if (gettimeofday(&tp2, &tzp) == -1) {
          perror("pitch time of day failed");
          exit(0);
     }
     if (pid)
          kill(pid, SIGUSR2);
/*
 * Close the socket
 */
     close(s);
     secs = tp2.tv_sec - tp1.tv_sec;
     usec = tp2.tv_usec - tp1.tv_usec;
     if (usec < 0) {
          secs;
          usec += 1000000;
     }
     msec = 1000*(float)secs;
     msec += (float)usec/1000;
     printf("PITCH: %d Kbytes/sec\n", (int)(bytes/msec));
     printf("CATCH: %d Kbytes/sec\n", bullet.throughput);
     printf("AVG:   %d Kbytes/sec\n",
 ((int)(bytes/msec)+bullet.throughput)/2);
}
```

7 Using UNIX Domain Datagram Sockets

This chapter describes communication between processes using UNIX Domain datagram sockets.

Overview

The UNIX Domain only allows communication between processes executing on the same machine. In contrast to pipes, it does not require the communicating processes to have common ancestry. For more information on the UNIX Domain protocol, refer to the `unix(7p)` man page.

UNIX domain (AF_UNIX) datagram sockets provide bidirectional, reliable, unduplicated flow of data while preserving record boundaries. Domain sockets significantly improve performance when compared to local IP loopback, due primarily to the lower code execution overhead and the fact that data is looped back at the protocol layer rather than at the driver layer.

AF_UNIX datagram sockets allow you to send and receive messages *without establishing a connection*. Each message includes a destination address. Processes involved in data transfer are not required to have a client-server relationship; the processes can be symmetrical.

AF_UNIX datagram sockets allow you to send to many destinations from one socket, and receive from many sources with one socket. There is no two-process model, although a two-process model is the simplest case of a more general multi-process model. The terms **server** and **client** are used in this section only in the application sense. For example, you might have a server process that receives requests from several clients on the same machine. This server process can send replies back to the various clients. This can all be done with one AF_UNIX datagram socket for the server.

The simplest two-process model is used in this section to describe AF_UNIX datagram sockets.

The following table lists the steps required to exchange data between AF_UNIX datagram sockets.

Table 7-1 **Exchanging Data Between UNIX Domain Datagram Sockets**

Client Process Activity	System Call Used	Server Process Activity	System Call Used
create a socket	`socket()`	create a socket	`socket()`
bind a socket	`bind()`	bind a socket	`bind()`
send message	`sendto()` or `sendmsg()`		
		receive message	`recvfrom()` or `recvmsg()`
		send message	`sendto()` or `sendmsg()`
receive message	`recvfrom()` or `recvmsg()`		

Each of these steps or activities is described in more detail in the following sections. The description of each activity specifies a system and includes:

- What happens when the system call is used.
- When to make the call.
- What the parameters do.
- Where to find details on the system call.

The domain datagram sockets programming examples are at the end of these descriptive sections. You can refer to the example code as you work through the descriptions.

Preparing Address Variables

Before your client process can make a request of the server process, you must establish the correct variables and collect the information you need about the server process.

Your server process needs to:

- Declare socket address variables.
- Get the pathname (character string) for the service you want to provide.

Your client process needs to:

- Declare socket address variables.
- Get the pathname (character string) for the service you want to use.

These activities are described next. Refer to the program example at the end of this chapter to see how these activities work together.

Declaring Socket Address Variables

You need to declare a variable of type struct `sockaddr_un` to use for the socket address for both processes. For example, the following declaration is used in the example server program:

```
struct sockaddr_un    servaddr;     /* server socket address */
```

`sockaddr_un` is a special case of `sockaddr` and is used with AF_UNIX address domain. The `sockaddr_un` address structure is defined in `sys/un.h` and consists of the following fields:

Field	Description
`short sun_family`	Specifies the address family and should always be set to AF_UNIX.
`u_char sun_path[92]`	Specifies the pathname to which the socket is bound or will be bound (e.g.: `/tmp/myserver`).

The server process only needs one address for its socket. Any process that knows the address of the server process can then send messages to it. Thus, your client process needs to know the address of the server socket. The client process will not need an address for its own socket, unless other processes need to refer to the client process.

Writing the Server and Client Processes

This section discusses the calls your server and client processes must make.

Creating Sockets

Both processes must call `socket` to create communication endpoints. `socket` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/socket.h>
```

System call:
```
s = socket(af, type, protocol)
int af, type, protocol;
```

Parameter	Description of Contents	INPUT Value
af	address family	AF_UNIX
type	socket type	SOCK_DGRAM
protocol	underlying protocol to be used	0 (default)

Function result: socket number (HP-UX file descriptor) if successful, -1 if socket call fails.

Example:
```
#include <sys/type.h>
#include <sys/socket.h>
...
s = socket(AF_UNIX, SOCK_DGRAM, 0)
```

When to Create Sockets

The server or client process should create sockets before any other BSD Sockets system calls. Refer to the `socket(2)` man page for more information on `socket`.

Binding Socket Addresses to UNIX Domain Datagram Sockets

After your server process has created a socket, it must call `bind` to bind a socket address. Until the server socket is bound to an address, other processes have no way to reference it.

The server process must bind a specific pathname to its socket. Set up the address structure with a local address before the server makes a call to `bind`.

The `bind` system call creates the inode file. If the inode file is not deallocated after bound sockets are closed, the names will continue to exist and cause directories to fill up with the unused files. To avoid directories filling up with these unused files, you can remove the files by calling `unlink` or remove them manually with the `rm` command. `bind` and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <sys/socket.h>
#include <sys/un.h>
```

System call:

```
bind(s, addr, addrlen);
int s;
struct sockaddr_un *addr;
int addrlen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket to be bound
addr	socket address	pointer to address to be bound to s
addrlen	length of socket address	size of struct sockaddr_un address

Function result: 0 if bind is successful, -1 if bind fails.

Example:

```
#include <sys/type.h>
#include <sys/socket.h>
#include <sys/un.h>
#define  SOCKET_PATH "/tmp/myserver"
struct  sockaddr_un     servaddr;
...
servaddr.sun_family = AF_UNIX;
```

```
strcpy(servaddr.sun_path, SOCKET_PATH);
unlink(SOCKET_PATH);
bind(s, &servaddr, sizeof(struct sockaddr_un));
```

When to Bind Socket Addresses

The server process should bind socket addresses after socket is created and before any other BSD Sockets system calls. Refer to the `bind(2)` man page for more information on `bind`.

Sending and Receiving Messages

The `sendto` and `recvfrom` (or `sendmsg` and `recvmsg`) system calls are usually used to transmit and receive messages with datagram sockets.

Sending Messages

Use `sendto` or `sendmsg` to send messages. `sendmsg` is similar to `sendto`, except `sendmsg` allows the send data to be gathered from several buffers. `sendto` and its parameters are described in the following table.

Include files:

```
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
```

System call:

```
count = sendto(s, msg, len, flags, to, tolen)
int s;
char *msg;
int len, flags;
struct sockaddr_un *to;
int tolen;
```

Parameter	Description of Contents	INPUT Value
s	socket descriptor of local socket	socket descriptor of socket that is sending the message
msg	pointer to data buffer	pointer to data to be sent
len	size of data buffer	size of msg
flags	settings for optional flags	0 (no options are currently supported)
to	address of recipient socket	pointer to the socket address that message should be sent to
tolen	size of to	length of address structure that to points to

Function result: number of bytes actually sent if sendto succeeds, -1 if sendto call fails.

Example:
```
struct  sockaddr_un     servaddr;
...
count = sendto(s, argv[2], strlen(argv[2]), 0,
&servaddr,
sizeof(struct sockaddr_un);
```

When to Send Data

The server or client process should send data after server has bound to an address. Refer to the `send(2)` man page for more information on `sendto` and `sendmsg`.

Receiving Messages

Use `recvfrom` or `recvmsg` to receive messages. `recvmsg` is similar to `recvfrom`, except `recvmsg` allows the read data to be scattered into buffers.

`recv` can also be used if you do not need to know what socket sent the message. However, if you want to send a response to the message, you must know where it came from. Except for the extra information returned by `recvfrom`, the two calls are identical. `recvfrom` and its parameters are described in the following table.

Include files:
```
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
```

System call:
```
count = recvfrom(s, msg, len, flags, from,
fromlen)
int s;
char *msg;
int len, flags;
struct sockaddr_un *from;
int *fromlen;
```

Parameter	Contents	INPUT Value	OUTPUT Value
s	socket descriptor of local socket	socket descriptor of socket receiving the message	unchanged
msg	pointer to data buffer	pointer to buffer that is to receive data	pointer to received data
len	maximum number of bytes that should be received	size of data buffer	unchanged
flags	settings for optional flags	0 (no options are supported	unchanged
from	address of socket that sent message	pointer to address structure, not used for input	pointer to socket address of socket that sent the message
fromlen	pointer to the size of from	pointer to size of from	pointer to the actual size of address returned

Function result: number of bytes actually received if recvfrom succeeds, -1 if recvfrom call fails.

Example:

```
struct  sockaddr_un     fromaddr;
int     fromlen;
...
count = recvfrom(s, msg, sizeof(msg), 0,
&fromaddr, &fromlen);
```

`recvfrom` blocks until there is a message to be received.

No more than `len` bytes of data are returned. The entire message is read in one `recvfrom`, `recvmsg`, `recv`, or `read` operation. If the message is too long for the receive buffer, the excess data are discarded. Because only one message can be returned in a `recvfrom` call, if a second

message is in the queue, it is not affected. Therefore, the best technique is to receive as much as possible on each call. Refer to the `recv(2)` man page for more information on `recvfrom` and `recvmsg`.

Closing a Socket

In most applications, you do not have to close the sockets. When you exit your program and your process terminates, the sockets are closed for you.

If you need to close a socket while your program is still running, use the `close` system call.

You may have more than one process with the same socket descriptor if the process with the socket descriptor executes a `fork`. `close` decrements the file descriptor count and the calling process can no longer use that file descriptor. When the last `close` is executed on a socket, any unsent messages are sent and the socket is closed. Any unreceived data are lost.

Example Using UNIX Domain Datagram Sockets

NOTE These programs are provided as examples only of UNIX Domain datagram socket usage and are not Hewlett-Packard supported products.

These programming examples demonstrate how to set up and use UNIX Domain datagram sockets. These sample programs can be found in the `/usr/lib/demos/networking/af_unix` directory. The client program is intended to run in conjunction with the server program.

This example shows how to create UNIX Domain datagram sockets and how to set up address structures for the sockets. In this example the client process sends 2000 bytes of data to the server (five times). The server process can receive data from any other process and will echo the data back to the sender.

The source code for these two programs follows.

```
/*
 *      AF_UNIX datagram server process
 *
 *      This is an example program that demonstrates the use of
 *      AF_UNIX datagram sockets as a BSD Sockets mechanism. This
 *      program contains the server and is intended to operate in
 *      conjunction with the client program.
 *
 */
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <sys/un.h>
#include <stdio.h>
#include <signal.h>
#include <netdb.h>

#define SOCKET_PATH     "/tmp/myserver"
#define bzero(ptr, len)  memset((ptr), NULL, (len))

int     timeout();
main()
{
        int     sock;
        int     slen, rlen, expect;
        unsigned char   sdata[5000];
        struct  sockaddr_un servaddr; /* address of server */
        struct  sockaddr_un from;
        int     fromlen;
        /* Escape hatch so blocking calls don't wait forever */
        signal(SIGALRM,timeout);
```

```
        alarm((unsigned long) 120);
        /*      Create a UNIX datagram socket for server        */
        if ((sock = socket(AF_UNIX, SOCK_DGRAM, 0)) < 0) {
                perror("server: socket");
                exit(1);
        }
        /*      Set up address structure for server socket      */

        bzero(&servaddr, sizeof(servaddr));
        servaddr.sun_family = AF_UNIX;
        strcpy(servaddr.sun_path, SOCKET_PATH);

        if (bind(sock, &servaddr, sizeof(servaddr)) < 0) {
                close(sock);
                perror("server: bind");
                exit(2);
        }
/*      Receive data from anyone, echo back data to the sender
 *      Note that fromlen is passed as pointer so recvfrom
 *      call can return the size of the returned address.
        */
        expect = 5 * 2000;
while (expect > 0) {
        fromlen = sizeof(from);
        rlen = recvfrom(sock, sdata, 2000, 0, &from, &fromlen);
        if (rlen == -1) {
               perror("server : recv\n");
               exit(3);
        } else {
              expect -= rlen;
              printf("server : recv'd %d bytes\n",rlen);
              slen = sendto(sock, sdata, rlen, 0, &from,
                            fromlen);
              if (slen <0) {
                       perror ("server : sendto\n");
                       exit (4);
                        }
                }
        }
        /*      Use unlink to remove the file (inode) so that the
        *       name will be available for the next run.
        */
        unlink(SOCKET_PATH);
        close(sock);
        printf("Server done\n");
        exit(0);
}
timeout() /* escape hatch so blocking calls don't wait forever */
{
        printf( "alarm received — stopping server\n" );
        fprintf(stderr, "stopping the server process\n");
        exit(5);
}
/*
*       AF_UNIX datagram client process
*
*       This is an example program that demonstrates the use of
*       AF_UNIX datagram sockets as a BSD Sockets mechanism.  This
*       contains the client, and is intended to operate in
*       conjunction with the server program.
```

```
 *
 */
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <sys/un.h>
#include <stdio.h>
#include <signal.h>
#include <netdb.h>
#define SOCKET_PATH     "/tmp/myserver"
#define SOCKET_PATHCLNT "/tmp/my_af_unix_client"
#define bzero(ptr, len)  memset((ptr), NULL, (len))
int     timeout();

main()
{
        int     sock;
        int     j, slen, rlen;
        unsigned char   sdata[2000];/* send data */
        unsigned char   rdata[2000];/* receive data */
        struct  sockaddr_un servaddr;/* address of server */
        struct  sockaddr_un clntaddr;/* address of client */
        struct  sockaddr_un from;
        int     fromlen;

        /*      Stop the program if not done in 2 minutes */

        signal(SIGALRM, timeout);
        alarm((unsigned long) 120);

        /* Fork the server process to receive data from client */

        printf("Client : Forking server\n");
        if (fork() == 0 ) {
                execl("./server", "server", 0 );
                printf("Cannot exec ./server.\n");
                exit(1);
        }
        /*      Initialize the send data        */

        for (j = 0; j < sizeof(sdata); j++)
                sdata[j] = (char) j;

        /* Create a UNIX datagram socket for client */

        if ((sock = socket(AF_UNIX, SOCK_DGRAM, 0)) < 0) {
                perror("client: socket");
                exit(2);
        }

        /*      Client will bind to an address so the server will
         *      get an address in its recvfrom call and use it to
         *      send data back to the client.
         */
        bzero(&clntaddr, sizeof(clntaddr));
        clntaddr.sun_family = AF_UNIX;
        strcpy(clntaddr.sun_path, SOCKET_PATHCLNT);

        if (bind(sock, &clntaddr, sizeof(clntaddr)) < 0) {
                close(sock);
```

```
                perror("client: bind");
                exit(3);
        }

        /*      Set up address structure for server socket */
        bzero(&servaddr, sizeof(servaddr));
        servaddr.sun_family = AF_UNIX;
        strcpy(servaddr.sun_path, SOCKET_PATH);

        for (j = 0; j < 5; j++) {
                sleep(1);
                slen = sendto(sock, sdata, 2000, 0,
                      (struct sockaddr *) &servaddr,
                        sizeof(servaddr));
                if (slen<0) {
                      perror("client: sendto");
                      exit(4);
                }
                else {
                     printf("client : sent %d bytes\n", slen);
                     fromlen = sizeof(from);
                     rlen = recvfrom(sock, rdata, 2000, 0, &from,
                                     &fromlen);
                     if (rlen == -1) {
                                perror("client: recvfrom\n");
                                exit(5);
                     } else
                     printf("client : received %d bytes\n", rlen);
                }
        }
        /*      Use unlink to remove the file (inode) so that the
         *      name will be available for the next run.
         */
        sleep(1);
        unlink(SOCKET_PATHCLNT);
        close(sock);
        printf("Client done\n");
        exit(0);
}
timeout() /* escape hatch so blocking calls don't wait forever */
{
        printf( "alarm went off - stopping client\n" );
        fprintf(stderr, "stopping the client process\n");
        exit(6);
}
```

8 Programming Hints

This chapter contains information for:

- Troubleshooting.
- Using diagnostic utilities as troubleshooting tools.

- Adding a server process to the internet daemon.
- Summary tables for system and library calls.
- Portability issues.

Troubleshooting

The first step to take is to avoid many problems by using good programming and debugging techniques. Your programs should check for a returned error after each system call and print any that occur. For example, the following program lines print an error message for `read`:

```
cc=read(sock,buffer,1000);
if (cc<0) {
      perror ("reading message")
      exit(1)
}
```

Refer to the `perror(3C)` man page for more information . Also refer to the appropriate man page for information about errors returned by the BSD Sockets system calls such as `read`.

You can also compile your program with the debugging option (-g) and use one of the debuggers (e.g. `cdb` or `xdb`) to help debug the programs.

It is possible that you could assign a reserved port address and cause a service to fail. For example, if the `nftdaemon` is not running, and you assign its port, when you try to start the `nftdaemon`, it fails. See the `/etc/services` file for the list of reserved ports.

Using Diagnostic Utilities as Troubleshooting Tools

You can use the following diagnostic utilities to help debug your programs. It is helpful if you have multiple access to the system so you can obtain information about the program while it is running.

`ping` — Use `ping` to verify the physical connection with the destination node.

`netstat` — Use `netstat` to display sockets and associations to help you troubleshoot problems in your application programs. Use `netstat` to determine if your program has successfully created a connection. If you are using stream sockets (TCP protocol), `netstat` can provide the TCP state of the connection. To check the status of a connection at any point in the program, use the `sleep` (seconds) statement in your program to pause the program. While the program is paused, execute `netstat -a` from another terminal.

Network Tracing — Network Tracing can be used to trace packets. For the trace information to be useful, you must have a working knowledge of network protocols.

Network Event Logging — Network Event Logging is an error logging mechanism. Use it in conjunction with other diagnostic tools.

These utilities are described in detail in the *Installing and Administering LAN/9000* manual.

Adding a Server Process to the Internet Daemon

This section contains example BSD Sockets programs that use the internet daemon, called `inetd`. For more information on `inetd`, refer to the `inetd(1M)` man page.

You can invoke the example server programs from `inetd` if you have super-user capabilities and you make the following configuration modifications:

- Add the following lines to the `/etc/inetd.conf` file:

```
example  stream  tcp  nowait  root  <path>/server.tcp  server.tcp
example   dgram   udp  wait    root    <path>/server.udp    server.udp
```

  where `<path>` is the path to the files on **your** host.

- Add the following lines to the `/etc/services` file:

```
example 22375/tcp
example 22375/udp
```

- If `inetd` is already running, execute the following command so that `inetd` recognizes the changes:

```
/etc/inetd -c
```

These example programs do the same thing as the previous example servers do, but they are designed to be called from `inetd`. They do not have daemon loops or listen for incoming connection requests, because `inetd` does that. The source code for the two example servers follows.

```
/*
 *                        S E R V E R . T C P
 *
 *      This is a variation of the example program called serv.tcp.
 *      This one performs the same function, except that it is
 *      designed to be called from /etc/inetd. This version does
 *      not contain a daemon loop, and does not listen for incoming
 *      connections on the socket. /etc/inetd does these functions.
 *      This server simply assumes that the socket to receive the
 *      messages from and send the responses to is file descriptor
 *      0 when the program is started.  It also assumes that
 *      the client connection is already established to the socket.
 *      For the sake of simplicity, the activity logging
 *      functions of serv.tcp have also been removed.
 *
 */

/*
```

```
 *                              M A I N
 *
 *  This is the actual server routine that the /etc/inetd forks to
 *  handle each individual connection.  Its purpose is to receive
 *  the request packets from the remote client, process them,
 *  and return the results to the client.
 *
 */
main()
{
    char buf[10];     /* This example uses 10 byte messages. */
     int len, len1;

        /* Go into a loop, receiving requests from the remote
         * client.  After the client has sent the last request,
         * it will do a shutdown for sending, which will cause
         * an end-of-file condition to appear on this end of the
         * connection.  After all the client's requests have
         * been received, the next recv call will return zero
         * bytes, signaling an end-of-file condition.  This is
         * how the server will know that no more requests will
         * follow, and the loop will be exited.
         */
        while (len = recv(0, buf, 10, 0)) {
                if (len == -1) {
                        exit (1); /* error from recv */
                }
          /* The reason this while loop exists is that there
           * is a remote possibility of the above recv returning
           * less than 10 bytes.  This is because a recv returns
           * as soon as there is some data, and will not wait for
           * all of the requested data to arrive.  Since 10 bytes
           * is relatively small compared to the allowed TCP
           * packet sizes, a partial receive is unlikely.  If
           * this example had used 2048 bytes requests instead,
           * a partial receive would be far more likely.
           * This loop will keep receiving until all 10 bytes
           * have been received, thus guaranteeing that the
           * next recv at the top of the loop will start at the
           * beginning of the next request.
           */
           while (len < 10) {
                   len1 = recv(0, &buf[len], 10-len, 0);
                   if (len1 == -1) {
                            exit (1);
                        }
                        len += len1;
                }
                         /* This sleep simulates the processing of
                          * the request that a real server may do.
                          */
                sleep(1);
                          /* Send a response back to the client. */
                if (send(0, buf, 10, 0) != 10) {
                        exit (1);
                }
        }
}
                /* The loop has terminated, because there are no
                 * more requests to be serviced.
```

```
        exit (0);
}
/*
 *                     S E R V E R . U D P
 *
 *      This is a variation of the example program called serv.udp.
 *      This one performs the same function, except that it is
 *      designed to be called from /etc/inetd. This version does
 *      not contain a daemon loop, and does not wait for requests
 *      to arrive on a socket. /etc/inetd does these functions. The
 *      server simply assumes the socket to receive the message
 *      from and send the response to is file descriptor 0 when
 *      the program is started.  It also assumes that the client's
 *      request is already ready to be received from the socket.
 *
 */
#include <sys/types.h>
#include <netinet/in.h>
#include <stdio.h>
#include <netdb.h>

#define BUFFERSIZE  1024  /* max size of packets to be received */
int cc;                   /* contains the number of bytes read */
char buffer[BUFFERSIZE];  /* buffer for packets to be read into */

struct hostent *hp;       /* pointer to info for requested host */

struct sockaddr_in clientaddr_in;/* for client's socket address */
struct in_addr reqaddr;          /* for requested host's address */

#define ADDRNOTFOUND  0xffffffff  /* return address for unfound
                                     host */

/*
 *                     M A I N
 *
 *      This routine receives the request and returns an answer.
 *      Each request consists of a host name for which the
 *      requester desires to know the internet address. The
 *      server will look up the name in its /etc/hosts file,
 *      and return the internet address to the client. An
 *      a internet address value of all ones will be returned
 *      if the host name is not found.
 *
 */
main()
{
        int  addrlen;

                 /* clear out address structure */
   memset ((char *)&clientaddr_in, 0, sizeof(struct sockaddr_in));

                /* Note that addrlen is passed as a pointer
                 * so that the recvfrom call can return the
                 * size of the returned address.
                 */
   addrlen = sizeof(struct sockaddr_in);
                /* This call will
                 * return the address of the client,
                 * and a buffer containing its request.
```

```
                 * BUFFERSIZE - 1 bytes are read so that
                 * room is left at the end of the buffer
                 * for a null character.
                 */
cc = recvfrom(0, buffer, BUFFERSIZE - 1, 0 &clientaddr_in,
                                                &addrlen);
        if ( cc == -1) exit(1);
                /* Make sure the message received in
                 * null terminated.
                 */
        buffer[cc]='\0';
                /* Treat the message as a string containing
                 * a hostname.  Search for the name
                 * in /etc/hosts.
                 */
        hp = gethostbyname (buffer);
        if (hp == NULL) {
                        /* Name was not found.  Return a
                         * special value signifying the error.
                        /*
                reqaddr.s_addr = ADDRNOTFOUND;
        } else {
                        /* Copy address of host into the
                         * return buffer.
                        /*
                reqaddr.s_addr =
                       ((struct in_addr *)(hp->h_addr))->s_addr;
        }
        /* send the response back to the requesting client.  The
         * address is sent in network byte order.  Note that
         * all errors are ignored.  The client
         * will retry if it does not receive
         * the response.
         */
        sendto (0, &reqaddr, sizeof(struct in_addr), 0,
                                &clientaddr_in, addrlen);
        exit(0);
}
```

Summary Tables for System and Library Calls

The following table contains a summary of the BSD Sockets system calls.

Table 8-1 **BSD Sockets System Calls**

System Call	Description
`socket`	Creates a socket, or communication endpoint for the calling process.
`bind`	Assigns a socket address to the socket specified by the calling process.
`listen`	Sets up a queue for incoming connection requests. (Stream sockets only.)
`connect`	For stream sockets, requests and creates a connection between the remote socket (specified by address) and the socket (specified by descriptor) of the calling process. For datagram sockets, permanently specifies the remote peer socket.
`accept`	Receives a connection between the socket of the calling process and the socket specified in the associated connect call. (Stream sockets only.)
`send`, `sendto`, `sendmsg`	Sends data from the specified socket.
`recv`, `recvfrom`, `recvmsg`	Receives data at the specified socket.
`shutdown`	Disconnects the specified socket.

System Call	Description
`getsockname`	Gets the socket address of the specified socket.
`getsockopt`, `setsockopt`	Gets, or sets, the options associated with a socket.
`getpeername`	Gets the name of the peer socket connected to the specified socket.

The following table contains a summary of the other system calls that can be used with BSD Sockets.

Table 8-2 Other BSD Sockets System Calls

System Call	Description
`read`	Can be used to read data at stream or datagram sockets just like recv or recvfrom , without the benefit of the recv flags. Read offers implementation independence; the descriptor can be for a file, a socket or any other object.
`write`	Can be used to write data from stream sockets (and datagram sockets if you declare a default remote socket address) just like send. Write offers implementation independence; the descriptor can be for a file, a socket or any other object.
`close`	Deallocates socket descriptors. The last close can be used to destroy a socket. Close does a graceful disconnect or a hard close, depending on the LINGER option. Refer to the sections on "Closing a Socket."

System Call	Description
`select`	Can be used to improve efficiency for a process that accesses multiple sockets or other I/O devices simultaneously. Refer to the sections on "Synchronous I/O Multiplexing with Select."
`ioctl`	Can be used for finding the number of receivable bytes with FIONREAD and for setting the nonblocking I/O flag with FIOSBNBIO. Can also be used for setting a socket to receive asynchronous signals with FIOASYNC .
`fcntl`	Can be used for duplicating a socket descriptor and for setting the O_NDELAY or O_NONBLOCK flag.

BSD Sockets attempts to isolate host-specific information from applications by providing library calls that return the necessary information.

The following table contains a summary of the library calls used with BSD Sockets. The library calls are in the common "c" library named `libc.sl`. Therefore, there is no need to specify any library name on the `cc` command line to use these library calls, `libc.sl` is used automatically.

Table 8-3 **Library Calls**

Library Call	Description
inet_addr inet_lnaof inet_makeaddr inet_netof inet_network	internet address manipulation routines
setservent endservent getservbyname getservbyport getservent	get or set service entry
setprotoent endprotoent getprotobyname getprotobynumber getprotoent	get or set protocol entry
setnetent endnetent getnetbyaddr getnetbyname getnetent	get or set network entry
sethostent endhostent gethostbyaddr gethostbyname gethostent	get or set host entry

Portability Issues

This section describes implementation differences between 4.3 BSD Sockets and HP-UX IPC. It contains porting issues for:

- IPC functions and library calls.
- Other functions and library calls typically used by IPC programs.

Because HP-UX IPC is based on 4.3 BSD Sockets (it is a subset of 4.3 BSD), programs should port easily between HP-UX and 4.3 BSD systems. If you need to have portable applications, keep the information described in this appendix in mind when you write your IPC programs.

Porting Issues for IPC Functions and Library Calls

The following is a list of differences in IPC functions and library calls to watch out for if you want to port your IPC applications between HP-UX and 4.3 BSD systems.

Shutdown

When `shutdown` has been used on a datagram socket on an HP-UX system, the local port number bound to that socket remains unavailable for use until that socket has been destroyed by close.

Some other systems free that port number for use immediately after the `shutdown`. In general, sockets should be destroyed by `close` (or by terminating the process) when they are no longer needed. This allows you to avoid unnecessary delay in deallocating local port numbers.

Address Conversion Functions for DEC VAX Hosts

The functions `htonl`, `htons`, `ntonl` and `ntons` are not required on HP-UX systems. They are included for porting to a DEC VAX host. You can use these functions in your HP-UX programs for portability; they are defined as null macros on HP-UX systems, and are found in `netinet/in.h`.

FIONREAD Return Values

For HP-UX systems, the FIONREAD `ioctl` request on a datagram socket returns a number that may be larger than the number of bytes actually readable. Previously, HP-UX systems returned the maximum number of bytes that a subsequent `recv` would be able to return.

Listen's Backlog Parameter

HP-UX sets the actual size of the queue for pending connections to `3/2*B+1`, where B is the `backlog` value specified in the `listen()` function. This may change in future releases, but it will not be smaller than the `backlog` value.

Pending Connections

There is no guarantee which pending connection on a listening socket is returned by `accept`. HP-UX systems return the newest pending connection. Applications should be written such that they do not depend upon connections being returned by `accept` on a first-come, first-served basis.

Porting Issues for Other Functions and Library Calls Typically Used by IPC

The following is a list of differences in functions and library calls to watch out for when you port your IPC applications between HP-UX and 4.3 BSD systems.

Ioctl and Fcntl Calls

4.3 BSD terminal `ioctl` calls are incompatible with the HP-UX implementation. These calls are typically used in virtual terminal applications. The HP-UX implementation uses UNIX System V compatible calls.

Pty Location

Look for the `pty` masters in `/dev/ptym/ptyp?` and for the `pty` slaves in `/dev/pty/ttyp?`. An alternative location to check is `/dev`.

Utmp

The 4.3 BSD `/etc/utmp` file format is incompatible with the HP-UX implementation. The HP-UX implementation uses UNIX System V compatible calls. Refer to the `utmp(4)` man page for details.

Library Equivalencies

Certain commonly used library calls in 4.3 BSD are not present in HP-UX systems, but they do have HP-UX equivalents. To make code porting easier, use the following equivalent library calls. You can do this by putting them in an include file, or by adding the define statements (listed in the following table) to your code.

Table 8-4 **Definition of Library Equivalents**

Define Statement	4.3 BSD Library	HP-UX Library
#define	index(a,b)	strchr(a,b)
#define	rindex(a,b)	strrchr(a,b)
#define	bcmp(a,b,c)	memcmp(a,b,c)
#define	bcopy(a,b,c)	memcpy(b,a,c)
#define	bzero(a,b)	memset(a,0,b)
#define	getwd(a)	getcwd(a,MAXPATHLEN)

NOTE Include `string.h` before using `strchr` and `strrchr`. Include `sys/param.h` before using `getcwd`.

Signal Calls

Normal HP-UX signal calls are different from 4.3 BSD signals. See the `sigvector(2)` man page for information on signal implementation.

Sprintf Return Value

For 4.3 BSD, `sprintf` returns a pointer to a string. For HP-UX systems, sprintf returns a count of the number of characters in the buffer.

A BSD Sockets Quick Reference Table

This appendix compares HP-UX BSD Sockets and X/Open Sockets.

Quick Reference Table

Release 10.10 of the HP-UX operating system supports two variations of sockets--HP-UX BSD Sockets and X/Open Sockets. To use X/Open sockets, users must make an addition to their make files by including the "-l xnet" argument with the "c89" or "cc" utilities. For detailed information about HP- UX BSD sockets, see this manual as well as the man pages. For detailed information about X/Open sockets, see "CAE Specification, Networking Services, Issue 4, X/Open" and the man pages. Note that the "msghdr" structures have different definitions between HP-UX BSD Sockets and X/ Open Sockets for the RECVMSG and SENDMSG calls. See the appropriate documentation for more information.

Table A-1 Comparison of HP-UX and X/Open Sockets

HP-UX	X/Open
accept int accept (int socket, void *address, int *address_len);	accept int accept (int socket, struct sockaddr *address, size_t *address_len);
bind int bind (int socket, const void *address, int address_len);	bind int bind (int socket, const struct sockaddr *address, size_t address_len);
close int close (int fildes);	close int close (int fildes);
connect int connect (int socket, const void *address, int *address_len);	connect int connect (int socket, const struct sockaddr *address, size_t address_len);
fcntl int fcntl (int fildes, int cmd, .../* arg */);	fcntl int fcntl (int fildes, int cmd, .../* arg */);

HP-UX	X/Open
getpeername int getpeername (int socket, void *address, int *address_len);	getpeername int getpeername (int socket, struct sockaddr *address, size_t *address_len);
getsockname int getsockname (int socket, void *address, int *address_len);	getsockname int getsockname (int socket, struct sockaddr *addr, size_t *address_len);
getsockopt int getsockopt (int socket, int level, int option_name, void *option_value, int *option_len);	getsockopt int getsockopt (int socket, int level, int option_name, void *option_value, int *option_len);
listen int listen (int s, int backlog);	listen int listen (int s, int backlog);
poll int poll (struct pollfd fds[], int nfds, int timeout);	poll int poll (struct pollfd fds[], int nfds, int timeout);
read(2) ssize_t read (int fildes, void *buf, size_t nbyte);	read(2) ssize_t read (int fildes, void *buf, size_t nbyte);
readv ssize_t readv (int fildes, const struct iovec *iov, int iovcnt);	readv ssize_t readv (int fildes, const struct iovec *iov, int iovcnt);
recv int recv (int socket, void *buffer, int length, int flags);	recv ssize_t recv (int socket, void *buffer, size_t length, int flags);
recvfrom int recvfrom (int socket, void *buffer, int length, int flags, void *address, int *address_len);	recvfrom ssize_t recvfrom (int socket, void *buffer, size_t length, int flags, struct sockaddr *address, size_t *address_len);

HP-UX	X/Open
recvmsg int recvmsg (int socket, struct msghdr msg[], int flags);	recvmsg ssize_t recvmsg (int socket, struct msghdr *msg, int flags);
select int select (int nfds, fd_set *readfds, fd_set *writefds, fd_set *errorfds, struct timeval *timeout);	select int select (int nfds, fd_set *readfds, fd_set *writefds, fd_set *errorfds, struct timeval *timeout);
send int send (int socket, const void *buffer, int length, int flags);	send ssize_t send (int socket, const void *buffer, size_t length, int flags);
sendmsg int sendmsg (int socket, const struct msghdr msg[], int flags);	sendmsg ssize_t sendmsg (int socket, const struct msghdr *msg, int flags);
sendto int sendto (int socket, const void *msg, int length, int flags, const void *dest_addr, int addr_len);	sendto ssize_t sendto (int socket, const void *msg, size_t length);
setsockopt int setsockopt (int socket, int level, int option_name, const void *option_value, int option_len);	setsockopt int setsockopt (int socket, int level, int option_name, const void *option_value, size_t option_len);
shutdown int shutdown (int s, int how);	shutdown int shutdown (int s, int how);
socket int socket (int af, int type, int protocol);	socket int socket (int domain, int type, int protocol);

HP-UX	X/Open
socketpair int socketpair (int af, int type, int protocol, int sv[2]);	socketpair int socketpair (int domain, int type, int protocol, int socketvector[2]);
write(2) ssize_t write (int fildes, const void *buf, size_t nbyte);	write(2) ssize_t write (int fildes, const void *buf, size_t nbyte);
writev ssize_t writev (int fildes, const struct iovec *iov, int iovcnt);	writev ssize_t writev (int fildes, const struct iovec *iov, int iovcnt);

Quick Reference Table

Glossary

Address family: The address format used to interpret addresses specified in socket operations. The internet address family (AF_INET) is supported.

Address: An Interprocess Communication term that refers to the means of labeling a socket so that it is distinguishable from other sockets, and routes to that socket are able to be determined.

Advanced Research Projects Agency: A U.S. government research agency that was instrumental in developing and using the original ARPA Services on the ARPANET.

Alias: A term used to refer to alternate names for networks, hosts and protocols. This is also an internetwork mailing term that refers an alternate name for a recipient or list of recipients (a mailing list).

ARPA: *See Advanced Research Projects Agency.*

ARPA/Berkeley Services:

The set of services originally developed for use on the ARPANET (i.e., telnet(1)) or distributed with the Berkeley Software Distribution of UNIX, version 4.3 (i.e., rlogin(1)).

ARPANET: The Advanced Research Projects Agency Network.

Association: An Interprocess Communication connection (e.g., a socket) is defined by an association. An association contains the (protocol, local address, local port, remote address, remote port)-tuple. Associations must be unique; duplicate associations on the same system may not exist.

Asynchronous Sockets: Sockets set up via ioctl with the FIOASYNC option to be notified with a SIGIO signal whenever a change on the socket occurs. Primarily used for sending and receiving data without blocking.

Berkeley Software Distribution: A version of

Glossary

UNIX software released by the University of California at Berkeley.

Binding: Establishing the address of a socket which allows other sockets to connect to it or to send data to it.

BSD: *See Berkeley Software Distribution.*

Channel: A communication path created by establishing a connection between sockets.

Client: A process that is requesting some service from another process.

Client host: The host on which a client process is running.

Communication domain: A set of properties that describes the characteristics of processes communicating through sockets. Only the Internet domain is supported.

Connection: A communications path to send and receive data. A connection is uniquely identified by the pair of sockets at either end of the connection.

See also, "Association."

Daemon: A software process that runs continuously and provides services on request.

DARPA: *See Defense Advanced Research Projects Agency.*

Datagram sockets: A socket that maintains record boundaries and treats data as individual messages rather than a stream of bytes. Messages may be sent to and received from many other datagram sockets. Datagram sockets do not support the concept of a connection. Messages could be lost or duplicated and may not arrive in the same sequence sent. Datagram sockets use the User Datagram Protocol.

Defense Advanced Research Projects Agency: The military arm of the Advanced Research Projects Agency. DARPA is instrumental in defining standards for ARPA

Glossary

services.

Domain: A set of allowable names or values. See also, "Communication domain."

File Transfer Protocol: The file transfer protocol that is traditionally used in ARPA networks. The ftp command uses the FTP protocol.

Forwarding: The process of forwarding a mail message to another destination (i.e., another user name, host name or network).

4.2 BSD: *See Berkeley Software Distribution.*

Frame: *See Packet.*

FTP: *See File Transfer Protocol.*

Gateway: A node that connects two or more networks together and routes packets between those networks.

Host: A node that has primary functions other than switching data for the network.

International Standards Organization: Called "ISO," this organization created a network model that identifies the seven commonly-used protocol levels for networking.

Internet: All ARPA networks that are registered with the Network Information Center.

Internet address: A four-byte quantity that is distinct from a link-level address and is the network address of a computer node. This address identifies both which network is on the Internet and which host is on the network.

Internetwork: A term used to mean "among different physical networks."

Interprocess Communication: A facility that allows a process to communicate with another process on the same host or on a remote host. IPC provides system calls that access sockets. This facility is distinct from Bell System V IPC. See also, "Sockets."

Glossary

IPC: *See Interprocess Communication.*

ISO: *See International Standards Organization.*

Link-level address: A six-byte quantity that is distinct from the internet address and is the unique address of the LAN interface card on each LAN.

Message: In IPC, the data sent in one UDP packet. When using sendmail a message is the information unit transferred by mail.

Node: A computer system that is attached to or is part of a computer network.

Node manager: The person who is responsible for managing the networking services on a specific node or host.

Official host name: The first host name in each entry in the /etc/hosts file. The official host name cannot be an alias.

Packet: A data unit that is transmitted between processes. Also called a "frame."

Peer: An Interprocess Communication socket at the other end of a connection.

Port: An address within a host that is used to differentiate between multiple sockets with the same internet address.

Protocol: A set of conventions for transferring information between computers on a network (e.g., UDP or TCP).

Remote host: A computer that is accessible through the network or via a gateway.

Reserved port: A port number between 1 and 1023 that is only for super-user use.

Server: A process or host that performs operations that local or remote client hosts request.

Service: A facility that uses Interprocess Communication to perform remote functions for a user (e.g., rlogin(1) or telnet(1)).

Glossary

Socket: Addressable entities that are at either end of an Interprocess Communication connection. A socket is identified by a socket descriptor. A program can write data to and read data from a socket, just as it writes and reads data to and from files.

Socket address: The internet address, port address and address family of a socket. The port and internet address combination allows the network to locate a socket.

Socket descriptor: An HP-UX file descriptor accessed for reading, writing or any standard file system calls after an Interprocess Communication connection is established. All Interprocess Communication system calls use socket descriptors as arguments.

Stream socket: A socket that, when connected to another stream socket, passes data as a byte stream (with no record boundaries). Data is guaranteed to arrive in the sequence sent. Stream sockets use the TCP protocol.

TCP: *See Transmission Control Protocol.*

Telnet: A virtual terminal protocol traditionally used on ARPA networks that allows a user to log into a remote host. The telnet command uses the Telnet protocol.

Transmission Control Protocol: A protocol that provides the underlying communication support for AF_INET stream sockets. TCP is used to implement reliable, sequenced, flow-controlled two-way communication based on a stream of bytes similar to pipes.

UDP: *See User Datagram Protocol.*

UNIX Domain Address: A character string containing the UNIX pathname to a UNIX Domain socket.

UNIX Domain Protocol: A protocol providing fast communication between

processes executing on the same node and using the AF_UNIX socket address family.

User Datagram Protocol: A protocol that provides the underlying communication support for datagram sockets. UDP is an unreliable protocol. A process receiving messages on a datagram socket could find that messages are duplicated, out-of-sequence or missing. Messages retain their record boundaries and are sent as individually addressed packets. There is no concept of a connection between the communicating sockets.

Virtual Terminal Protocol: A protocol that provides terminal access to interactive services on remote hosts (e.g., telnet(1)).

Index

Index

Index

Index

EXHIBIT F

FULLY REDACTED